BERND LEINER

Ökonometrische Schätzverfahren

— Eine kritische Übersicht —

Saarbrücken 1969

Ökonometrische Schätzverfahren

- Eine kritische Übersicht -

I n a u g u r a l d i s s e r t a t i o n

zur Erlangung des akademischen Grades

eines

Doktors der Wirtschaftswissenschaft

der

Rechts- und Wirtschaftswissenschaftlichen Fakultät der Universität des Saarlandes

(doctor rerum politicarum)

vorgelegt von:

Diplom-Volkswirt

Bernd Leiner

UNIV.
LEIDEN
BIBL.

Tag der Disputation: 24.7.1969

Dekan: Prof. Dr. H. Zacher

Erstgutachter: Prof. Dr. G. Menges

Zweitgutachter: Prof. Dr. H. Schreeweiß

Meiner Frau Maja

Einleitung 1
ERSTER TEIL: Darstellung 5
1. Die Voraussetzungen 6
2. Die gewöhnliche Methode der kleinsten Quadrate ... 9
3. Die Eigenschaften der gewöhnlichen Methode der kleinsten Quadrate 22
4. Die verallgemeinerte Methode der kleinsten Quadrate 35
5. Die zweistufige Methode der kleinsten Quadrate ... 43
6. Die dreistufige Methode der kleinsten Quadrate ... 51
7. Die indirekte Methode der kleinsten Quadrate 58
8. Die Methode der größten Dichte bei vollständiger Information 60
9. Die Methode der größten Dichte bei beschränkter Information nach Anderson und Rubin 71
10. Die Methode der größten Dichte bei beschränkter Information nach Koopmans und Hood 91
11. Das Prinzip des minimalen Streuungsverhältnisses 109
12. Die Schätzungen der k-Klasse 115
13. Three-Pass Least Squares 125
14. Die Momentenmethode 131
15. Die χ^2-Minimum-Methode 134
16. Die Methode der Hilfsvariablen 137
17. Sonstige Schätzmethoden 139
ZWEITER TEIL: Kritischer Vergleich 141
18. Abgrenzung 141
19. Die Eigenschaften guter Schätzungen 143
20. Eine einzige Gleichung mit einer endogenen Variablen 145
21. Ein reduzierte-Form-Gleichungssystem, das keine verzögerten endogenen Variablen enthält 153
22. Ein reduzierte-Form-Gleichungssystem, das verzögerte endogene Variablen enthält 164

23. Eine einzige Gleichung mit mehr als einer unverzögerten endogenen Variablen 169
24. Ein interdependentes Gleichungssystem 177
DRITTER TEIL: Synthese 189
25. Ergebnisse des kritischen Teils 189
26. Allgemeine Probleme der Anwendbarkeit ökonometrischer Schätzverfahren auf ökonomische Tatbestände 200
27. Stellungnahme zu den Monte-Carlo-Studien 202
28. Probleme beim Auftreten von Autokorrelation 206
29. Schlußbetrachtung 212
LITERATURVERZEICHNIS 213

Einleitung

Die Ökonometrie, deren Geburtsjahr[1] das Jahr 1930 ist, kann als eine gemeinsame Grenzwissenschaft[2] der Disziplinen Mathematik, Statistik und Nationalökonomie betrachtet werden. Somit ist es nicht verwunderlich, daß eine solche moderne und aus derart heterogenen Wurzeln stammende Wissenschaft sich noch heutzutage in einem ausgeprägten Methodenstreit befindet. Während der Statistik als Wissenschaft noch eine Vermittlerstellung zwischen den Wissenschaften Mathematik und Nationalökonomie zuerkannt werden kann, treffen im Rahmen der ökonometrischen Wissenschaft mathematische Methoden und nationalökonomische Betrachtungen hart aufeinander. Diese Erkenntnis drängt sich insbesondere auf, wenn man die Entwicklung der ökonometrischen Forschungsweise kritisch betrachtet. Am offenkundigsten gilt dies für die Anwendungsweise des ökonometrischen Instrumentariums, welches hauptsächlich aus den ökonometrischen Schätzverfahren besteht.

Anfänglich stand im wesentlichen den ökonometrischen Forschern als leicht berechenbare und plausible ökonometrische Schätzmethode nur die Methode der kleinsten Quadrate zur Verfügung, von der sie in den zu jener Zeit üblichen mikroökonomischen Partialanalysen Gebrauch machten. Daß diese Methode zu jener Zeit oft kritiklos angewandt wurde, gibt sogar Herman Wold, einer der stärksten Verfechter der ökonometrischen Anwendbarkeit der Methode der kleinsten Quadrate, zu:[3] "Die meisten der früheren Autoren

---

1) Vgl. Menges, G.: Ökonometrie. Wiesbaden 1961, S. 9.

2) Vgl. Menges, G.: Die Unterschiede in den Erkenntnisweisen von Statistik und Ökonometrie. Annales Universitatis Saraviensis (Rechts- und Wirtschaftswissenschaften), 7 (1959), S. 191-202, s. bes. S. 193.
Vgl. Menges (1961): a.a.O., S. 16.

3) Wold, H. (in Zusammenarbeit mit Juréen, L.): Demand Analysis. New York-Stockholm 1953, S. 28-29. Übersetzung des Verfassers.

waren nicht sehr ausführlich bezüglich des Rationalen der Regressionsmethoden. Nicht, daß irgendwelche tiefgründigen Beweise zur Interpretation und logischen Fundierung verlangt werden. Eher werden die grundlegenden Streitfragen im allgemeinen behandelt, als würden sie eine ausführliche Behandlung nicht nötig haben, wobei die Beweise beiläufig oder überhaupt nicht übermittelt werden, so daß sie durch Lesen zwischen den Zeilen abgeleitet werden müssen."

Erst Trygve Haavelmo[1)] machte im Jahre 1943 auf die Verzerrung der Schätzungen aufmerksam, die entsteht, sobald die Methode der kleinsten Quadrate in unbedachter Weise auf die Gleichungen eines simultanen Gleichungssystems angewandt wird.

Dies kann als Auftakt des noch heute andauernden Methodenstreites angesehen werden, denn in der Folgezeit wurden die ökonometrischen Voraussetzungen der bisherigen Schätzverfahren mathematisch fundiert und die Schätzergebnisse mittels einer Reihe von Kriterien überprüft. Als Ergebnisse der Prämissen- und Schätzresultatskritik präsentierten sich nunmehr Schätzverfahren, die zwar mathematisch komplizierter wurden, aber deren Voraussetzungen verallgemeinerter oder realitätsnäher waren und/oder deren Schätzresultate den angelegten Kriterien besser entsprachen.

Berücksichtigt man, daß ökonometrische Methoden auf Gleichungen und Gleichungssysteme der verschiedensten Art angewendet werden können, so ist es nicht verwunderlich, daß die Brauchbarkeit jeder der zur Zeit bekannten ökono-

---

1) Haavelmo, T.: The statistical implications of a system of simultaneous equations. Econometrica, 11(1943), S. 1-12.

metrischen Schätzmethoden je nach Voraussetzung und Anwendung stark divergieren kann. Da die gegenwärtige Diskussion ökonometrischer Schätzmethoden jeweils nur einige Aspekte dieses vielfältigen Komplexes beleuchten kann, hat sich der Verfasser dieser Arbeit die Mühe gemacht, die wichtigsten heute bekannten ökonometrischen Schätzverfahren darzustellen und durch kritischen und systematischen Vergleich der Voraussetzungen, Eigenschaften und Anwendungsmöglichkeiten dieser Verfahren neue Aspekte der ökonometrischen Schätzdiskussion zu erarbeiten.

Die vorliegende Arbeit zerfällt in einen Teil, der sich der Darstellung widmet, und einen Teil, der sich mit der aus dieser Darstellung folgenden Kritik befaßt. Abschliessend faßt der Teil "Synthese" die wichtigsten Erkenntnisse, die aus dieser Arbeit resultieren, zusammen.

Die Darstellung und Kritik der modernen ökonometrischen Schätzverfahren, insbesondere hinsichtlich der Anwendung auf simultane Gleichungssysteme, ist dermaßen kompliziert und vielschichtig, daß eine andere Präsentation als in Matrizenschreibweise schwerfällig und unübersichtlich wäre. Infolgedessen fühlte sich der Verfasser veranlaßt, sich die Matrizenschreibweise[1)] als eine Art mathematische

---

1) Folgende Publikationen wurden zur Erlernung der Matrizenschreibweise herangezogen:

Aitken, A.C.: Determinants and Matrices. Edinburgh-London 1939 (9. Aufl. 1958).

Zurmühl, R.: Matrizen und ihre technischen Anwendungen. Berlin-Göttingen-Heidelberg 1964.

Johnston, J.: Econometric Methods. New York-San Francisco-Toronto-London 1963, Chapter 3.

Menges (1961): a.a.O., S. 220-229.

Im folgenden werden allgemeine Rechenregeln der Matrizenrechnung nicht zitiert, da sie in diesen Publikationen nachgelesen werden können.

Kurzschrift anzueignen, um die neueren Entwicklungen der Schätztheorie verfolgen und interpretieren zu können. Während zu Beginn dieser Arbeit nur Kenntnis des Aussehens einer Matrix und eines Vektors sowie der elementweisen Addition, Subtraktion und besonders Multiplikation sowie einige Rechenregeln der Kehrmatrix und der transponierten Matrix vorausgesetzt werden, nimmt später der Schwierigkeitsgrad allmählich zu, wobei jedoch nach Möglichkeit erklärende Anmerkungen die Übergänge allgemein verständlich machen sollen.

E R S T E R T E I L

Darstellung

In diesem Teil soll versucht werden, die wichtigsten Schätzverfahren, die zum Teil recht kompliziert sind, in aller Ausführlichkeit darzustellen. Diese Ausführlichkeit führt spätestens im zweiten Teil dieser Arbeit zu einem Gewinn, wenn beim kritischen Vergleich verschiedener Schätzmethoden miteinander Zweifel über die Voraussetzungen, das Vorgehen und die Eigenschaften einer oder mehrerer Schätzmethoden auftauchen. Nach Möglichkeit wurden zur Darstellung diejenigen Publikationen herangezogen, in denen das betrachtete Schätzverfahren zum ersten Male präsentiert wurde. Hiervon wurde abgewichen, wenn das Verfahren später in moderner Terminologie oder mit leichter vergleichbaren Symbolen von einem anderen Autor übernommen wurde. Waren die verwendeten Symbole allzu irreführend, so wurde das Verfahren mit anderen Symbolen formuliert. Letzteres erwies sich jedoch als so mühsam, daß sonst der leichteren Vergleichbarkeit mit den Originalbeiträgen der Vorzug gegeben wurde. So wurde diese Arbeit nicht in einer durchgehenden Symbolik geschrieben, was jedoch die Vergleichbarkeit der einzelnen Schätzergebnisse miteinander materiell nicht beeinflussen konnte.

Bevor die einzelnen Schätzmethoden im einzelnen dargestellt werden, sollen die wichtigsten Voraussetzungen, die von den einzelnen Methoden in unterschiedlicher Kombination verwandt werden, summarisch dargestellt werden. Gerade für das Verständnis und die Kritik ökonometrischer Schätzverfahren ist eine genaue Kenntnis der Voraussetzungen unumgänglich, da diese die Schätzergebnisse und die Anwendbarkeit der Schätzverfahren auf ökonomische Gegebenheiten in eindeutiger Weise bedingen.

1. Die Voraussetzungen[1)]

Während in der ökonomischen Struktur

$$(1.1) \qquad f_i(y_t, z_t) = 0 \qquad (i=1,\ldots,n)$$

nur die direkt meßbaren Variablen enthalten sind, enthält die ökonometrische Struktur

$$(1.2) \qquad f_i(y_t, z_t) = u_{it} \qquad (i=1,\ldots,n)$$

darüberhinaus die nicht direkt meßbaren Variablen u_{it}, d.h. die latenten Variablen. Die latenten Variablen üben also einen nicht direkt meßbaren Einfluß auf die Struktur aus, da die Spezifikation der Vielzahl der Einflußfaktoren nicht möglich oder aus Gründen der Identifikation[2)] nicht praktikabel ist und damit werden über den Einfluß der nicht im Modell spezifizierten Variablen gewisse Annahmen gemacht. Der Einfluß aller sogenannten Störvariablen, die in der i-ten Gleichung für die Perioden t=1,...,T nicht spezifiziert sind, wird also in der latenten Variablen u_{it} zusammengefaßt.

Voraussetzung 1:
Die latenten Variablen u_{it} werden als Zufallsvariablen aufgefaßt mit dem Erwartungswert Null.

$$(1.3) \qquad E(u_{it}) = 0.$$

In dieser Voraussetzung kommt zum Ausdruck, daß die latente Variable eine stochastische Größe ist, die bestimmten Zufallsgesetzen gehorcht und den Mittelwert Null hat. Der Einfluß aller latenten Variablen zusammen kann also als

---

1) Vgl. hierzu Menges (1961): a.a.O., S. 60-66.
Menges, G.: Three Essays in Econometrics. Statistische Hefte, 4(1963), S. 1-37, s. bes. S. 2-10.

2) Menges, G.: Die ökonometrische Struktur und die Frage ihrer Konstanz. In: Strukturwandlungen einer wachsenden Wirtschaft. Bd. 2, Berlin 1964 (Schriften des Vereins für Socialpolitik, Neue Folge, Bd. 30), S. 989-1004, s. bes. S. 995-996.

unwesentlich erwartet werden. Über alle Gleichungen und Zeiträume des Modells betrachtet, gleichen sich also die entgegengerichteten Einflüsse, die in der latenten Variablen zusammengefaßt sind, gegeneinander aus.

Für den Spaltenvektor $u' = \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix}$ gilt dann also $E(u') = 0$,

wobei 0 der T-zeilige Nullvektor ist.

Voraussetzung 2:

Die latenten Variablen u_{it} sind unabhängig von allen vorherbestimmten Variablen des Modells. Daher gilt[1)]

(1.4) $\quad \text{cov}\,(u_t, z_{t-\theta}) = 0 \quad$ für alle t und θ.

Voraussetzung 3:

Die Streuungs-Kovarianz-Matrix $\mathfrak{m}$ der latenten Variablen hat folgende Gestalt

(1.5) $$\mathfrak{m} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{bmatrix}$$

Die σ_{ij} mit $i \neq j$ sind die Kovarianzen $\text{cov}\,(u_{it}, u_{jt})$.

Die σ_{ij} mit $i=j$ sind die Streuungen $\text{cov}\,(u_{it}, u_{it}) = \sigma^2_{u_i}$

Wir nehmen nun an, daß $\mathfrak{m}$ regulär ist, d.h., daß

(1.6) $\quad \det \mathfrak{m} \neq 0.$

---

1) Valavanis, S.: Econometrics. An Introduction to Maximum Likelihood Methods. New York-Toronto-London 1959 (posthum hrsgg. von A.H. Conrad), S. 18.

Voraussetzung 4:
Die u_t für t=1,...,T sind stochastisch unabhängig voneinander. Aus dieser Annahme resultiert

$$(1.7) \qquad p(u) = \prod_{t=1}^{T} p_t(u_t).$$

Voraussetzung 5:
Die u_t für t=1,...,T gehorchen demselben Verteilungsgesetz, d.h. das Verteilungsgesetz der u_t ist zeitlich konstant in dem Zeitraum von t=1 bis t=T. Es gilt also

$$(1.8) \qquad p_t(u_t) = p(u) \qquad \text{für } t=1,\ldots,T.$$

Voraussetzung 6:
Die u_{it} für i=1,...,n und festes t sind voneinander unabhängig. Es gilt also

$$(1.9) \qquad E(u_{it}, u_{jt}) = 0 \qquad \text{für } i,j=1,\ldots,n, \text{ und t fest.}$$

Voraussetzung 7:
Wird die Unabhängigkeit dagegen nur für den Unterfall $i \neq j$ gefordert, dann gilt für die Kovarianzen

$$(1.10) \qquad E(u_{it}, u_{jt}) = 0 \qquad \text{für } i,j=1,\ldots,n, \text{ t fest und } i \neq j$$

Da die Streuungs-Kovarianz-Matrix hierbei nur noch auf der Hauptdiagonalen von Null verschiedenen Elementen aufweist, also folgende Gestalt hat

$$(1.11) \qquad \mathfrak{m} = \begin{bmatrix} \sigma_{11} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \sigma_{nn} \end{bmatrix},$$

nennt man dies den Diagonalfall.

Voraussetzung 7 ist daher schwächer als Voraussetzung 6.

Voraussetzung 8:
Die u_{it} sind bei festem t multinormal verteilt mit Streuungs-Kovarianz-Matrix $\mathfrak{m}$. Es gilt dann

$$(1.12) \quad p(u) = \frac{1}{\sqrt{(2\pi)^n (\det \mathfrak{m})}} \cdot \exp\left(-\frac{1}{2} u\, \mathfrak{m}^{-1}\, u'\right)$$

wobei $\mathfrak{m}^{-1}$ die Kehrmatrix von $\mathfrak{m}$ ist und u' der Spaltenvektor ist, den man durch das Transponieren des Zeilenvektors u erhält. Der in der Exponentialfunktion stehende Ausdruck lautet im einzelnen

$$(1.13) \quad u\, \mathfrak{m}^{-1}\, u' = \sum_{i=1}^{n} \sum_{j=1}^{n} u_i\, \sigma^{ij}\, u_j$$

wobei σ^{ij} die Elemente der inversen Matrix $\mathfrak{m}^{-1}$ sind. Beim Diagonalfall reduziert sich (1.12) zu

$$(1.14) \quad p(u) = \frac{1}{\sqrt{(2\pi)^n \prod_{i=1}^{n} \sigma_{ii}}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \frac{u_i^2}{\sigma_{ii}}\right).$$

2. Die gewöhnliche Methode der kleinsten Quadrate[1)]

Die gewöhnliche Methode der kleinsten Quadrate ist die älteste der hier behandelten Schätzmethoden. Sie

1) Vgl. Johnston: a.a.O., S. 9-12.

wurde[1] im Jahre 1795 von Carl Friedrich Gauß[2] entwickelt, um - nach verschiedenen geodätischen Anwendungen - aufgrund der fehlerbehafteten Beobachtungen der Planeten deren Umlaufbahn zu schätzen. So berechnete Gauß als erster die genauen Umlaufbahnen der Planeten Ceres und Pallas mittels der gewöhnlichen Methode der kleinsten Quadrate. Diese Methode wollen wir nur deshalb "gewöhnlich" nennen, weil sie sonst mit komplizierteren Varianten dieser Methode verwechselt werden könnte.

Bevor die etwas schwierige Matrizendarstellung herangezogen werden soll, erscheint es der besseren Einarbeit halber vorteilhafter, die Methode der kleinsten Quadrate zunächst mit Hilfe von Skalaren darzustellen.

Gegeben seien je n Beobachtungen oder Realisationen einer erklärten Variablen y und einer diese erklärenden Variablen x, nämlich

$$y_1, y_2, \cdots, y_n$$
$$x_1, x_2, \cdots, x_n.$$

Nehmen wir an, daß die erklärende Variable x die erklärte Variable y mittels einer linearen Funktion erklärt, so müßte diese lineare Funktion folgendes Aussehen haben

$$y = \alpha + \beta x \tag{2.1}$$

---

1) Vgl. hierzu Eisenhart, C.: The background and evolution of the method of least squares. Digest prepared for distribution to participants of the 34th Session of the International Statistical Institute, Ottawa, Canada, 21-29 August 1963, S. 16.

2) Herman Wold (in Zusammenarbeit mit L. Juréen) (a.a.O., S. 29) schreibt die Entdeckung der Methode fälschlich Legendre (1806) zu. Vgl. hierzu jedoch Eisenhart: a.a.O., S. 16, wonach Gauß bereits 1802 Olbers über die Methode schrieb, dieses Skript jedoch erst 1805 zurückbekam.

wobei α und β Koeffizienten, auch Parameter genannt, sind, und zwar α der Ordinatenabschnitt und β die Steigung. Diese Parameter sollen zeitlich konstant sein[1].

Wollen wir nun für alternierende (x_i/y_i)-Wertepaare (wobei $i=1,\ldots,n$) die Lage der zugehörigen linearen Funktion erfahren, so müssen wir die Werte der Parameter α und β, die uns bislang unbekannt sind, auf irgendeine Art und Weise schätzen. Die Notwendigkeit der Schätzung ergibt sich daraus, daß

1) im zweidimensionalen Raum die Lage einer Geraden (als Abbildung einer linearen Funktion) schon durch die Kenntnis der Koordinaten zweier Punkte, durch die diese Gerade verlaufen soll, bestimmt ist, so daß die Kenntnis jedes weiteren (x_i/y_i)-Wertepaares entweder überflüssig ist oder der Aussage, daß die beiden ersten Punkte auf der Geraden liegen, widerspricht,

2) wie im Gauß'schen Beispiel die einzelnen Beobachtungen durch zufällige Fehler (Meßfehler, Erhebungsfehler) gestört sind und daher um die gedachte Gerade streuen.

In aller Regel ist es nun sinnvoll, jede der Beobachtungen als Information in den Schätzvorgang einzubeziehen[2], d.h. wir versuchen, eine Gerade dermaßen durch den Punktehaufen $P_i(x_i/y_i)$ mit $i=1,\ldots,n$ zu legen, daß die quadratischen

---

1) Vgl. Menges, G. u. Diehl, H.: Das Stabilitätsproblem in der Ökonometrie. Statistische Hefte, 6(1965), S. 27-52, s. bes. S. 33-39 und Eigenschaft E_9 auf S. 33-34.

2) Bestehen a-priori-Kenntnisse darüber, daß eine oder mehrere der Beobachtungen nicht typisch für die zu schätzende Funktion sind, so wird ein solcher "Ausreißer" zur Schätzung nicht herangezogen.

Abstände der Beobachtungspunkte $P_i(x_i/y_i)$ von den zugehörigen Punkten auf der Schätzgeraden $\hat{P}_i(x_i/\hat{y}_i)$ eine minimale Summe bilden. Im Idealfall liegen die Beobachtungspunkte alle auf der Schätzgeraden, so daß $y_i = \hat{y}_i$ und damit $P_i = \hat{P}_i$ und die Summe der quadratischen Abstände gleich Null ist.

Die etwas schwierigen quadratischen Abstände werden den einfachen Abständen vorgezogen, weil die positiven, über der Geraden liegenden, Abstände $(y_i > \hat{y}_i)$ sonst bei der Summierung die negativen Abstände $(y_i < \hat{y}_i)$ kompensieren und eine Minimierung der Summe stets zur Summe Null führt, obwohl die Beobachtungswerte weit um die Schätzgerade streuen können. Die Berechnung quadratischer Abstände hat gerade den Vorteil, daß die am weitesten von der Schätzgeraden liegenden Beobachtungspunkte am stärksten in der Summe vertreten sind durch ihr quadratisches Gewicht und somit die Glaubwürdigkeit der Schätzgeraden auf die Probe stellen können, da aus ihnen die Streuung der Beobachtungswerte um die Schätzgerade berechnet werden kann.

Die quadratischen Abstände der Punkte P_i von den Punkten $\hat{P}_i$ auf der Schätzgeraden ergeben sich als quadratische Differenz der y_i-Werte von den $\hat{y}_i$-Werten, also der Ordinatenwerte, da die Abszissen x_i bei beiden Punkten gleich sind. Bezeichnen wir die Abstände mit e_i, so erhalten wir für die Summe der quadratischen Abstände

$$(2.2) \qquad \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Wir bezeichnen nun die Schätzgerade der linearen Funktion (2.1) mit

$$(2.3) \qquad \hat{y} = \hat{\alpha} + \hat{\beta}\, x_i,$$

wobei $\hat{\alpha}$ und $\hat{\beta}$ die nach der gewöhnlichen Methode der kleinsten Quadrate vorzunehmenden Schätzungen der Parameter α und β sein sollen.

Wenn wir (2.3) in (2.2) einsetzen, so erhalten wir

$$(2.4) \qquad \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta}\, x_i)^2 .$$

Wir suchen nun diejenigen Parameterschätzungen $\hat{\alpha}$ und $\hat{\beta}$, für die die Summe der Quadrate der Abweichungen $\sum_{i=1}^{n} e_i^2$ ein Minimum ist. Um dieses Minimum zu finden, müssen wir (2.4) partiell differenzieren bezüglich $\hat{\alpha}$ und $\hat{\beta}$ und erhalten die beiden Normalgleichungen

$$(2.5) \qquad \frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{\alpha}} = -2 \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta}\, x_i)$$

und

$$(2.6) \qquad \frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{\beta}} = -2 \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta}\, x_i) \cdot x_i$$

Wir können (2.5) und (2.6) Null gleichsetzen, weil im Minimum die erste Ableitung gleich Null sein muß. Dann erhalten wir, wenn wir zudem (2.5) durch (-2n) dividieren,

$$(2.7) \qquad \bar{y} = \hat{\alpha} + \hat{\beta}\, \bar{x}$$

oder

$$(2.8) \qquad \hat{\alpha} = \bar{y} - \hat{\beta}\, \bar{x},$$

wobei $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ bzw. $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ die Mittelwerte der y_i bzw. der x_i sind.

Entsprechend erhalten wir nach Nullsetzen und Division mit (-2n) aus (2.6)

$$(2.9) \qquad \frac{1}{n} \Sigma y_i x_i = \hat{\alpha} \bar{x} + \hat{\beta} \cdot \frac{1}{n} \Sigma x_i^2$$

Durch Einsetzen von (2.8) in (2.9) gelangt man zu

$$(2.10) \qquad \frac{1}{n} \Sigma y_i x_i - \bar{y} \bar{x} = \hat{\beta}\left(\frac{1}{n} \Sigma x_i^2 - \bar{x}^2\right)$$

und schließlich zu

$$(2.11) \qquad \hat{\beta} = \frac{\frac{1}{n} \Sigma y_i x_i - \bar{y} \bar{x}}{\frac{1}{n} \Sigma x_i^2 - \bar{x}^2} .$$

Die Schätzung der Steigung ß nach der gewöhnlichen Methode der kleinsten Quadrate ist das Verhältnis der Kovarianz zwischen x und y zur Streuung von x. Wir können (2.11) alternativ schreiben[1)]

$$(2.12) \qquad \hat{\beta} = \frac{\frac{1}{n} \Sigma(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n} \Sigma(x_i - \bar{x})^2} = \frac{\mathrm{cov}(x, y)}{\sigma_x^2} .$$

---

1) Es gilt nämlich

$$\frac{1}{n} \Sigma(x_i - \bar{x})^2 = \frac{1}{n} \Sigma(x_i - \bar{x})(x_i - \bar{x}) = \frac{1}{n} \Sigma x_i^2 - \bar{x}\bar{x} - \bar{x}\bar{x} + \bar{x}^2$$

$$= \frac{1}{n} \Sigma x_i^2 - \bar{x}^2$$

und

$$\frac{1}{n} \Sigma(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n} \Sigma x_i y_i - \bar{x}\bar{y} - \bar{y}\bar{x} + \bar{x}\bar{y} = \frac{1}{n}\Sigma y_i x_i - \bar{y}\bar{x}.$$

Nach der Methode der kleinsten Quadrate erhalten wir also die Schätzung für ß, nämlich $\hat{ß}$, durch Einsetzen der Beobachtungswerte und ihrer Mittelwerte in die Formel (2.12). Danach erhalten wir die Schätzung für α, nämlich $\hat{\alpha}$, durch Einsetzen von $\hat{ß}$ in die Formel (2.8).

Eine wichtige Eigenschaft der gewöhnlichen Methode der kleinsten Quadrate sollte noch festgestellt werden: Wenn wir in Gleichung (2.6) für den Klammerausdruck wieder e_i setzen, wie in Gleichung (2.4), und dann bedenken, daß (2.6) als erste Ableitung Null gleich sein muß, so erhalten wir nach Division mit (-2n)

$$(2.13) \qquad \frac{1}{n} \sum_{i=1}^{n} (e_i \, x_i) = 0.$$

Die Summe der Produkte der erklärenden Variablen x_i mit den zugehörigen, nach der Methode der kleinsten Quadrate ermittelten Residuen e_i ist gleich Null.

Wir wollen nun von der Einfachregression zur Mehrfachregression übergehen, d.h. wir wollen die erklärte Variable y in Abhängigkeit von mehreren erklärenden Variablen $x_1, \ldots, x_k$ schätzen. Da wir für jede Variable n Beobachtungswerte zur Verfügung haben, wenden wir der besseren Übersichtlichkeit halber nunmehr die Matrizenrechnung an.

Wir gehen von der folgenden Strukturgleichung aus[1)]

$$(2.14) \qquad y = x_1 ß_1 + x_2 ß_2 + \ldots + x_k ß_k + u$$

---

1) Vgl. im folgenden Johnston: a.a.O., S. 106-109.

Wieder sind die "wahren" Parameterwerte $\beta_1, \beta_2, \ldots, \beta_k$ unbekannt und müssen mit Hilfe der Beobachtungswerte und einer linearen Schätzgleichung geschätzt werden. u ist die latente Variable der Strukturgleichung und repräsentiert den nicht meßbaren Einfluß einer Gesamtheit von Störvariablen. Ein absolutes Glied wie in Gleichung (2.1) erscheint nicht explizit in Gleichung (2.14). Dennoch wird auch der inhomogene Fall durch Gleichung (2.14) hinreichend erfaßt, wenn[1)]

(a) die Variablen als Abweichungen ihres Mittelwerts gemessen werden, da bei der Subtraktion das absolute Glied verschwindet, oder

(b) eine der erklärenden Variablen eine Scheinvariable ist, die über alle Perioden den Wert 1 annimmt, so daß der zugehörige Parameter das absolute Glied ist.

Könnten wir die Strukturgleichung (2.14) über n Perioden beobachten, so erhielten wir folgendes System von Gleichungen[2)]

$$
\begin{aligned}
y_1 &= x_{11}\beta_1 + x_{21}\beta_2 + \ldots + x_{k1}\beta_k + u_1 \\
y_2 &= x_{12}\beta_1 + x_{22}\beta_2 + \ldots + x_{k2}\beta_k + u_2 \\
&\vdots \quad\; \vdots \quad\; \vdots \qquad\qquad \vdots \quad\; \vdots \\
y_n &= x_{1n}\beta_1 + x_{2n}\beta_2 + \ldots + x_{kn}\beta_k + u_n \,.
\end{aligned}
\tag{2.15}
$$

Mit der Matrizenschreibweise läßt sich (2.15) vereinfachen zu

---

1) Siehe Schneeweiß, H.: Vorlesung Ökonometrie, SS 1966, an der Universität des Saarlandes (hektographiert), S. 38.

2) Der Index n gibt die Zeit an und wird daher nachgestellt, obwohl er die Zeile kennzeichnet.

(2.16) $$\mathfrak{y} = \mathfrak{X}\,\beta + u,$$

wobei

(2.17) $$\mathfrak{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}, \quad \mathfrak{X} = \begin{bmatrix} x_{11} x_{21} & \cdots & x_{k1} \\ x_{12} x_{22} & \cdots & x_{k2} \\ \cdot \quad \cdot & & \cdot \\ \cdot \quad \cdot & & \cdot \\ x_{1n} x_{2n} & \cdots & x_{kn} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} \text{ und } u = \begin{bmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ \cdot \\ u_n \end{bmatrix}.$$

Wenn wir in (2.16) β, den Parametervektor, schätzen wollen, müssen wir folgende Voraussetzungen als erfüllt annehmen:

(2.18) $E(u) = 0$

(2.19) $E(u\,u') = \sigma^2 I_n$

(2.20) $\mathfrak{X}$ enthält nur fest vorgegebene Größen

(2.21) Der Rang von $\mathfrak{X}$ ist k, wobei $k < n$.

Die ersten beiden Voraussetzungen sind wohlbekannt[1] und besagen, daß der Erwartungswert der latenten Variablen gleich dem Nullvektor der Ordnung (n x 1) ist und die Streuungs-Kovarianz-Matrix der latenten Variablen eine Diagonalmatrix ist, wobei die Elemente der Hauptdiagonalen einander gleich sind. Dieser Spezialfall einer Diagonalmatrix wird auch Skalarmatrix genannt[2]. Da die Kovarianzen alle gleich Null sind, ist eine Autokorrelation der latenten Variablen, d.h. eine Korrelation der latenten Variablen verschiedener Perioden miteinander, ausgeschlossen. Da die Streuungen auf der Hauptdiagonalen in verschie-

1) Vgl. Kapitel 1.

2) Vgl. Zurmühl: a.a.O., S. 11-12.

denen Perioden identisch sind, spricht man auch von Homoskedastizität. Die in (2.20) formulierte Voraussetzung gewährleistet gemeinsam mit (2.18), daß die erklärenden Variablen mit den latenten Variablen nicht korreliert sind, denn (2.20) impliziert $E(\mathfrak{X}) = \mathfrak{X}$, so daß $E(\mathfrak{u}'\mathfrak{X}) = 0$. Der Rang von $\mathfrak{X}$ soll kleiner als n sein, damit mehr Beobachtungswerte als zu schätzende Parameter vorliegen. Dies ist eine fundamentale Voraussetzung, da sonst das Gleichungssystem (2.15) unterbestimmt wäre und die Schätzung aller Parameter unmöglich wäre.

Wir wollen nun die gewöhnliche Methode der kleinsten Quadrate auf (2.16) anwenden, also die Schätzung

$$(2.22) \qquad \hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{bmatrix}$$

des Parametervektors β ermitteln.

Wir bilden wieder[1] die Summe der Quadrate der Abweichungen e_i

$$(2.23) \qquad \sum_{i=1}^{n} e_i^2 = e'e = (\mathfrak{y} - \hat{\mathfrak{y}})'(\mathfrak{y} - \hat{\mathfrak{y}}) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 .$$

1) Vgl. (2.2).

wobei

$$(2.24) \qquad e = \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{bmatrix} \quad \text{und} \quad \hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{y}_n \end{bmatrix} .$$

Die lineare Schätzfunktion[1] lautet

$$(2.25) \qquad \hat{y} = X \hat{\beta}.$$

Durch Einsetzen von (2.25) in (2.23) erhalten wir[2]

$$(2.26) \qquad \begin{aligned} e'e &= (y - X\hat{\beta})'(y - X\hat{\beta}) \\ &= y'y - y'X\hat{\beta} - \hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta} \\ &= y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}, \end{aligned}$$

weil $y'X\hat{\beta}$ ein Skalarprodukt um eine Matrix X und daher ein Skalar[3] ist und der transponierte Skalar dem Skalar gleich ist, mithin $y'X\hat{\beta} = (y'X\hat{\beta})' = \hat{\beta}'X'y$.

Um die minimale Summe der Quadrate der Abweichungen zu finden, müssen wir (2.26) nach dem Schätzparametervektor

---

1) Vgl. entsprechend (2.3).

2) Vgl. auch Johnston: a.a.O., S. 108.

3) Natürlich ist auch bei Mehrfachregression innerhalb einer Gleichung die Summe der Quadrate der Abweichungen als Skalar noch die Summe bzw. Differenz von Skalaren. So sind $y'y$ und $\hat{\beta}'X'X\hat{\beta}$ ebenfalls Skalare.

$\hat{\beta}$ differenzieren. Die sicherste aber umständliche Art der Differentiation ist die elementweise Differentiation. Die Differentiation nach einem Vektor konnte ich in den herkömmlichen Lehrbüchern der Matrizenrechnung nicht finden. Gleichwohl möchte ich folgende Rechenregel anbieten: Wird ein Zeilenvektor der Länge k nach dem transponierten Vektor (dies ist der Spaltenvektor der Länge k) differenziert, so ergibt sich[1)] die Einheitsmatrix der Ordnung k.

---

1) Die elementweise Differentiation ergibt z.B. bei $\hat{\beta}'$ folgende Spalteneinheitsvektoren mit k Elementen:

$$\frac{\partial \hat{\beta}'}{\partial \hat{\beta}_1} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \frac{\partial \hat{\beta}'}{\partial \hat{\beta}_2} = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad \frac{\partial \hat{\beta}'}{\partial \hat{\beta}_k} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix},$$

was wir zusammenfassen können zu

$$\frac{\partial \hat{\beta}'}{\partial \hat{\beta}} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix} = I_k.$$

Diejenigen Elemente in (2.26), in denen $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ nicht enthalten sind, wie z.B. $x'y$ und $x'x$, sind für die partielle Differentiation als konstant anzusehen und können als Konstante behandelt werden. Somit ergibt sich für

$$\frac{\partial \hat{\beta}' x'b}{\partial \hat{\beta}} = I_k x' y = x' b$$

und für

$$\frac{\partial \hat{\beta}' x' x \hat{\beta}}{\partial \hat{\beta}} = 2 \cdot I_k x' x \hat{\beta} = 2 \cdot x' x \hat{\beta},$$

weil bei quadratischen Formen nur ein Vektor durchdifferenziert werden muß. [Fortsetzung auf S. 21]

Unter Berücksichtigung dieser Rechenregel erhalten wir für die 1. Ableitung von (2.26) nach $\hat{\beta}$

$$(2.27) \qquad \frac{\partial\, e'e}{\partial\, \hat{\beta}} = -2X'y + 2X'X\hat{\beta}.$$

Da im Minimum die 1. Ableitung Null gleichgesetzt werden kann, erhalten wir, wenn wir (2.27) durch 2 dividieren:

$$(2.28) \qquad X'X\hat{\beta} = X'y.$$

Durch Multiplikation mit $(X'X)^{-1}$ von links erhalten wir wegen $(X'X)^{-1}(X'X) = I_k$

$$(2.29) \qquad \hat{\beta} = (X'X)^{-1}X'y.$$

Dies ist die Schätzung des Parametervektors β, nämlich $\hat{\beta}$, nach der gewöhnlichen Methode der kleinsten Quadrate.

Setzen wir diese Schätzung von β, also $\hat{\beta}$, in die lineare Schätzfunktion (2.25) ein, so erhalten wir

$$(2.30) \qquad \hat{y} = X(X'X)^{-1}X'y$$

die lineare Schätzung von y.

Es sei hier noch bemerkt, daß die Differentiation eines Spaltenvektors der Länge k nach dem transponierten Vektor (dies ist der Zeilenvektor der Länge k) eine Einheitsmatrix der Ordnung k ergibt, die sich jedoch aus k Zeileneinheitsvektoren zusammensetzt. Außerdem wird darauf aufmerksam gemacht, daß die Differentiation eines Zeilenvektors nach sich selbst (also auch einem Zeilenvektor) nicht die Einheitsmatrix ergibt. Dasselbe gilt für die Differentiation eines Spaltenvektors nach sich selbst (also auch einen Spaltenvektor).

Verwenden wir die in (2.23) enthaltene Definition des Vektors der Abweichungen, so erhalten wir für

$$(2.31) \qquad \begin{aligned} e &= \mathfrak{y} - \hat{\mathfrak{y}} \\ &= \mathfrak{y} - \mathfrak{X}(\mathfrak{X}'\mathfrak{X})^{-1}\,\mathfrak{X}'\mathfrak{y} \\ &= [I_n - \mathfrak{X}(\mathfrak{X}'\mathfrak{X})^{-1}\,\mathfrak{X}']\,\mathfrak{y}. \end{aligned}$$

3. Die Eigenschaften der gewöhnlichen Methode der kleinsten Quadrate[1)2)]

Durch Einsetzen von (2.16) in (2.29) erhalten wir

$$(3.1) \qquad \hat{\mathfrak{B}} = (X'X)^{-1}\,X'(X\,\mathfrak{B} + \mathfrak{u})$$

oder

$$(3.2) \qquad \hat{\mathfrak{B}} = \mathfrak{B} + (X'X)^{-1}\,X'\mathfrak{u},$$

weil

$$(3.3) \qquad (X'X)^{-1}\,X'X = I_K \quad \text{ist.}$$

Der Erwartungswert von $\hat{\mathfrak{B}}$ ist dann

$$(3.4) \qquad E(\hat{\mathfrak{B}}) = E(\mathfrak{B}) + E[(X'X)^{-1}\,X'\mathfrak{u}].$$

---

1) Vgl. Johnston: a.a.O., S. 109-114.

2) Aus schreibtechnischen Gründen können die in diesem Kapitel vorkommenden Vektoren und Matrizen nicht besonders gekennzeichnet werden. Eine Verwechslungsmöglichkeit mit Skalaren besteht jedoch nicht, weil die außer den Skalaren σ^2, 2, n und K in diesem Kapitel vorkommenden Skalare besonders erwähnt werden. Die Definition der Vektoren und Matrizen sind im vorhergehenden Kapitel nachzulesen, soweit sie nicht in diesem Kapitel definiert sind.

Hieraus erhalten wir

$$(3.5.) \qquad E(\hat{\beta}) = \beta + (X'X)^{-1} E(X'u),$$

weil X nach (2.20) fest vorgegeben und der Erwartungswert des "wahren" β das "wahre" β selbst ist.

Sind nun u und X stochastisch unabhängig, so gelangen wir von (3.5) zu

$$(3.6) \qquad E(\hat{\beta}) = \beta,$$

d.h. $\hat{\beta}$ ist eine erwartungstreue Schätzung von β unter den gegebenen Voraussetzungen.

Die Parametervektorschätzung $\hat{\beta}$ mit der gewöhnlichen Methode der kleinsten Quadrate ist also erwartungstreu, wenn die exogenen Variablen X von den latenten Variablen u stochastisch unabhängig sind.

Wir berechnen nun die Streuungs-Kovarianz-Matrix der Schätzung des Parametervektors, welche wir mit dem Symbol SKP bezeichnen.

$$(3.7) \qquad SKP\ (\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'].$$

Da nach (3.2) gilt

$$(3.8) \qquad (\hat{\beta} - \beta) = (X'X)^{-1} X'u,$$

ist

$$(3.9) \qquad (\hat{\beta} - \beta)' = u'X(X'X)^{-1}$$

und wir erhalten für SKP $(\hat{\beta})$

$$(3.10) \qquad \text{SKP}\ (\hat{\beta}) = E[(X'X)^{-1}\ X'u\ u'X(X'X)^{-1}].$$

Mit Hilfe von Voraussetzungen (2.19) und (2.20) erhalten wir

$$(3.11) \qquad \text{SKP}\ (\hat{\beta}) = (X'X)^{-1}\ X'\sigma^2\ X(X'X)^{-1}$$

Da σ^2 ein Skalar ist und daher vorgezogen werden kann und dann $(X'X)^{-1}\ X'X = I_K$ gilt, gelangen wir zu dem Ergebnis

$$(3.12) \qquad \text{SKP}\ (\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = \sigma^2(X'X)^{-1}.$$

Wir wollen nun beweisen, daß unter den gegebenen Voraussetzungen die Methode der kleinsten Quadrate unter allen erwartungstreuen linearen Schätzungen von β in bezug auf γ die minimale Summe der Varianzen der Schätzfehler aufweist.[1)]

Zu diesem Zweck betrachten wir eine beliebige in γ lineare Vektorschätzung[2)] b von β

$$(3.13) \qquad b = \mathfrak{A}\ \gamma$$

wobei $\mathfrak{A}$ eine aus festen Werten bestehende Matrix der Ordnung (k x n) sei. Dann gilt $E(\mathfrak{A}) = \mathfrak{A}$. Damit b eine erwartungstreue Schätzung des "wahren" Parametervektors ist, muß gelten $E(b) = \beta$. Es ist aber

---

1) Es wird hier darauf verzichtet, die Minimumeigenschaft für jede einzelne Fehlervarianz zu zeigen. Eine entsprechende Darstellung findet sich bei Johnston: a.a.O., S. 16-19.

2) Auch dieser Vektor b muß wie β von der Ordnung (K x 1) sein.

(3.14) $E(b) = E(\mathfrak{A}\ y) = E[\mathfrak{A}(X\ \beta + u)]$ wegen (2.16)

und wir erhalten wegen $E(\mathfrak{A}\ u) = \mathfrak{A}[E(u)] = 0_n$ und $E(\mathfrak{A}\ X\ \beta) = \mathfrak{A}\ X\ \beta$

(3.15) $E(b) = \mathfrak{A}\ X\ \beta$

b ist demnach dann und nur dann erwartungstreu, d.h. $E(b) = \beta$, wenn

(3.16) $\mathfrak{A}\ X = I_k$.

Die Streuung der Schätzung b um β können wir angeben mit

(3.17) $E[(b - \beta)'(b - \beta)]$

und wir erhalten wegen (3.13) und (2.16) für

(3.18) $(b - \beta) = [\mathfrak{A}(X\ \beta + u) - \beta] = \mathfrak{A}\ X\ \beta - \beta + \mathfrak{A}\ u.$

Wenn (3.16) erfüllt ist, erhalten wir hieraus

(3.19) $(b - \beta) = \mathfrak{A}\ u$.

Aus (3.19) erhalten wir für $(b - \beta)' = u'\mathfrak{A}'$. Dann wird aus (3.17)

(3.20) $E[(b - \beta)'(b - \beta)] = E(u'\mathfrak{A}'\mathfrak{A}\ u)$.

$u'\mathfrak{A}'\mathfrak{A}\ u$ ist eine quadratische Form, wobei $\mathfrak{A}'\mathfrak{A}$ von einem Skalarprodukt umgeben ist, so daß $u'\mathfrak{A}'\mathfrak{A}\ u$ ein Skalar ist.

Da ein Skalar seiner Spur gleich ist (die Spur enthält dann nur ein Element, den Skalar selbst), können wir für (3.20) schreiben, wobei das Symbol Sp für Spur steht,

$$(3.21) \qquad E[(b - \beta)'(b - \beta)] = E[Sp(u'\mathfrak{A}'\mathfrak{A} u)]$$

Da $u'\mathfrak{A}'$ von der Ordnung $(1 \times k)$ und $\mathfrak{A} u$ vom Typ $(k \times 1)$ sind, können wir[1] innerhalb der Spur umgruppieren und es gilt

$$(3.22) \qquad \begin{aligned} E[Sp(u'\mathfrak{A}'\mathfrak{A} u)] &= E[Sp(\mathfrak{A} u u \mathfrak{A}')] \\ &= Sp[E(\mathfrak{A} u u'\mathfrak{A}')] \\ &= Sp\{\mathfrak{A}[E(u u')] \mathfrak{A}'\} \end{aligned}$$

und wir erhalten wegen (2.19)

$$(3.23) \qquad E[(b - \beta)'(b - \beta)] = \sigma^2 \, Sp(\mathfrak{A} \mathfrak{A}'),$$

weil ein Skalar auch vor die Spur gezogen werden kann.

Die Streuung von b bezüglich $\mathfrak{A}$ muß nun eine minimale sein unter der Restriktion (3.16), die die Erwartungstreue bedingt.

Wir verwenden zu dieser Optimierungsaufgabe Lagrange-Multiplikatoren und bilden die zu minimierende Funktion Φ

$$(3.24) \qquad \Phi = \sigma^2 \, Sp(\mathfrak{A} \mathfrak{A}') - Sp[\mathfrak{M}(\mathfrak{A} X - I_k)']$$

wobei $\mathfrak{M}$, die Matrix der Lagrange-Multiplikatoren, folgendes Aussehen hat

$$(3.25) \qquad \mathfrak{M} = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1k} \\ \vdots & & \vdots \\ \mu_{k1} & \cdots & \mu_{kk} \end{bmatrix}$$

1) Vgl. Menges (1961): a.a.O., S. 224.

Die Matrix $\mathfrak{M}$ besteht also aus k^2 Skalaren μ_{ij} wobei $i,j=1,\ldots,k$.

Wir differenzieren jetzt (3.24) bezüglich $\mathfrak{B}$ und erhalten

$$(3.26) \qquad \frac{\partial \Phi}{\partial \mathfrak{B}} = \sigma^2 \, Sp\left(\frac{\partial \mathfrak{B}\,\mathfrak{B}'}{\partial \mathfrak{B}}\right) - Sp\left(\frac{\partial \mathfrak{M}\, X'\mathfrak{B}'}{\partial \mathfrak{B}}\right).$$

Hier müssen wir wieder eine technische Betrachtung einschieben. Das Ergebnis der Ableitung einer Matrix der Ordnung $(n \times k)$ nach ihrer transponierten Matrix der Ordnung $(k \times n)$ konnte ich ebenfalls[1] in keinem der herkömmlichen Lehrbücher der Matrizenrechnung finden. Gleichwohl lautet das Ergebnis: Die Einheitsmatrix der Ordnung $(n\,k)$.[2] Wir betrachten dies im einzelnen.

Die Matrix $\mathfrak{B}'$ der Ordnung $(n \times k)$ bestehe aus folgenden Skalaren:

$$(3.27) \qquad \mathfrak{B}' = \begin{bmatrix} a_{11} & \cdots & a_{k1} \\ \vdots & & \vdots \\ a_{1n} & \cdots & a_{kn} \end{bmatrix} = [a_1 \ \cdots \ a_k],$$

$$\text{wobei } a_1 = \begin{bmatrix} a_{11} \\ \vdots \\ a_{1n} \end{bmatrix}, \ \ldots, \ a_k = \begin{bmatrix} a_{k1} \\ \vdots \\ a_{kn} \end{bmatrix}.$$

Wir können also $\mathfrak{B}'$ in k Spaltenvektoren a_i $(i=1,\ldots,k)$ zerlegen. Dann ergibt die Differentiation jedes Spalten-

---

1) Vgl. die Betrachtungen zur Berechnung von (2.27).
2) Genauer $[n \times (k \times k) \times n]$.

vektors der Ordnung (n x 1) nach seinem transponierten Vektor [dem Zeilenvektor der Ordnung (1 x n)] die Einheitsmatrix der Ordnung (n x n)[1]. Wir erhalten somit κ dieser Einheitsmatrizen I_n.

Dagegen ergibt die Differentiation des ersten Spaltenvektors $\begin{bmatrix} a_{11} \\ \vdots \\ a_{1n} \end{bmatrix}$ nach dem transponierten Vektor eines anderen Spaltenvektors, z.B. des letzten Spaltenvektors $\begin{bmatrix} a_{\kappa 1} \\ \vdots \\ a_{\kappa n} \end{bmatrix}$, stets die Nullmatrix der Ordnung n, also O_n, weil kein Element beiden Vektoren gemeinsam ist.

Führen wir die Differentiation systematisch durch, so müssen wir jede Spalte der Matrix $\mathfrak{A}'$ im Zähler nach jeder Zeile der Matrix $\mathfrak{A}$ im Nenner des Differentialbruches differenzieren.

Schematisch verwenden wir auch die Zerlegung von

$$(3.28) \qquad \mathfrak{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{\kappa 1} & \cdots & a_{\kappa n} \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_\kappa \end{bmatrix}$$

zur Differentiation:

---

1) Vgl. die Betrachtungen zur Berechnung von (2.27).

$$(3.29)\qquad \frac{\partial \pi'}{\partial \pi} = \frac{\partial [a_1\ a_2\ a_3\ \dots\ a_K]}{\partial \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_K \end{bmatrix}}$$

$$= \begin{bmatrix} I_n & 0_n & 0_n & \dots & 0_n \\ 0_n & I_n & 0_n & \dots & 0_n \\ 0_n & 0_n & I_n & \dots & 0_n \\ \vdots & \vdots & \vdots & & \\ 0_n & 0_n & 0_n & \dots & I_n \end{bmatrix} = I_{nk}$$

Jede der in I_{nk} enthaltenen (Einheits- oder Null-)Matrizen der Ordnung n muß schließlich noch mit den bei der Differentiation konstanten Matrizen, die zuvor mit π' multiplikativ verknüpft waren, multipliziert werden.

Dann erhalten wir für (3.26)

$$(3.30)\qquad \frac{\partial \phi}{\partial \pi} = 2\sigma^2\ \mathrm{Sp} \begin{bmatrix} \Sigma & 0_n & 0_n & \dots & 0_n \\ 0_n & \Sigma & 0_n & \dots & 0_n \\ 0_n & 0_n & \Sigma & \dots & 0_n \\ \vdots & \vdots & \vdots & & \vdots \\ 0_n & 0_n & 0_n & \dots & \Sigma \end{bmatrix} - \mathrm{Sp} \begin{bmatrix} \pi X' & 0_n & 0_n & \dots & 0_n \\ 0_n & \pi X' & 0_n & \dots & 0_n \\ 0_n & 0_n & \pi X' & \dots & 0_n \\ \vdots & \vdots & \vdots & & \vdots \\ 0_n & 0_n & 0_n & \dots & \pi X' \end{bmatrix}$$

$$= 2k \cdot \sigma^2 \cdot \Sigma - k \cdot \pi X',$$

da die Spur bekanntlich die Summe der (hier k identischen) Hauptdiagonalelemente einer Matrix ist.[1)]

1) Vgl. Menges (1961): a.a.O., S. 224.

Berücksichtigen wir, daß die erste Ableitung im Minimum gleich Null ist, hier also gleich der Nullmatrix von der Ordnung (k x n), so erhalten wir aus (3.30) nach Division durch den Skalar k

$$(3.31) \qquad 2\sigma^2 \mathfrak{U} = \mathfrak{M} X'.$$

Multiplizieren wir (3.31) von rechts mit X, so erhalten wir

$$(3.32) \qquad 2\sigma^2 \mathfrak{U} X = \mathfrak{M} X'X$$

und wegen $\mathfrak{U} X = I_k$ und nach Multiplikationen mit $(X'X)^{-1}$ von rechts

$$(3.33) \qquad \mathfrak{M} = 2\sigma^2 (X'X)^{-1}.$$

Durch Einsetzen von (3.33) in (3.31) gelangen wir zu

$$(3.34) \qquad \mathfrak{U} = (X'X)^{-1} X'$$

und, wenn wir (3.34) in (3.13) einsetzen, zu

$$(3.35) \qquad b = (X'X)^{-1} X'\mathfrak{y} = \hat{\beta}.$$

Da dies zugleich die Schätzung $\hat{\beta}$ nach der gewöhnlichen Methode der kleinsten Quadrate ist, ist bewiesen, daß unter den getroffenen Voraussetzungen die gewöhnliche Methode der kleinsten Quadrate unter allen erwartungstreuen linearen Schätzungen die minimale Summe der Varianzen der Schätzfehler aufweist.

Wir wollen[1)] nun prüfen, ob σ_e^2 eine erwartungstreue Schätzung von σ_u^2 ist, d.h. ob

---

1) Siehe Johnston: a.a.O., S. 112.

(3.36) $\quad E(e'e) \stackrel{?}{=} E(u'u).$

Wegen (2.26) haben wir

(3.37) $\quad e = y - X\hat{\beta}$

und wegen (2.16) und (2.29)

(3.38) $\quad e = X\beta + u - X[(X'X)^{-1} X'(X\beta + u)].$

Durch Ausmultiplizieren und Subtrahieren erhalten wir

(3.39) $\quad e = u - X(X'X)^{-1} X'u,$

woraus sich

(3.40) $\quad e = [I_n - X(X'X)^{-1} X']u$

bilden läßt.

e ist also eine lineare Funktion von u. Hierbei zeigt sich, daß der Ausdruck in den eckigen Klammern auf der rechten Seite von (3.40) eine symmetrische, idempotente[1] Matrix ist, denn es gilt

(3.41) $\quad [I_n - X(X'X)^{-1} X']' = [I_n - X(X'X)^{-1} X']$

und

(3.42) $\quad [I_n - X(X'X)^{-1}X'][I_n - X(X'X)^{-1}X'] = [I_n - X(X'X)^{-1}X']$

wegen $\quad I_n - 2X(X'X)^{-1}X' + X(X'X)^{-1}(X'X)(X'X)^{-1}X'$

$\quad = I_n - X(X'X)^{-1}X'.$

---

1) Vgl. Zurmühl: a.a.O., S. 114.

Dann erhalten wir für e'e aus (3.40)

$$(3.43) \qquad e'e = u'[I_n - X(X'X)^{-1}X']'[I_n - X(X'X)^{-1}X']\,u$$

und hieraus wegen (3.41) und (3.42)

$$(3.44) \qquad e'e = u'[I_n - X(X'X)^{-1}X']\,u$$

Betrachten wir nun e'e als skalares Produkt, nämlich als $\sum_{i=1}^{n} e_i e_i$ in Skalarschreibweise, wobei die e_i Skalare sind, so sehen wir, daß dies zugleich die Spur von e e' ist, denn e e' besteht aus folgenden Skalaren

$$(3.45) \qquad e\,e' = \begin{bmatrix} e_1 e_1 & \cdots & e_1 e_n \\ \vdots & & \vdots \\ e_n e_1 & \cdots & e_n e_n \end{bmatrix}$$

und hat die Spur

$$(3.46) \qquad Sp(e\,e') = \sum_{i=1}^{n} e_i e_i = e'e = Sp(e'e),$$

wobei letztere Gleichheit sich aus der Identität zwischen einem Skalar und seiner Spur ergibt.

Wir können nun (3.44) wegen (3.46) und (3.40) umformen zu

$$(3.47) \qquad \begin{aligned} e'e = Sp(e\,e') &= Sp\{[I_n - X(X'X)^{-1}X']uu' \\ &\qquad [I_n - X(X'X)^{-1}X']'\} \\ &= Sp\{u\,u'[I_n - X(X'X)^{-1}X']\}. \end{aligned}$$

Dann lautet der Erwartungswert von e'e

$$(3.48) \quad E(e'e) = E\ Sp\ \{u\,u'[I_n - X(X'X)^{-1}X']\}$$
$$= Sp\ E\ \{u\,u'[I_n - X(X'X)^{-1}X']\}$$
$$= Sp\ \{\sigma^2[I_n - X(X'X)^{-1}X']\}$$

wegen (2.19) und (2.20) und weil natürlich auch die Einheitsmatrix immer fest vorgegeben ist und daher ihrem Erwartungswert gleich ist.

Wir können jetzt weiter umbilden und gelangen zu

$$(3.49) \quad E(e'e) = \sigma^2\ Sp\ [I_n - X(X'X)^{-1}X']$$

weil σ^2 als Skalar vor die Spur gezogen werden kann.
Aus (3.49) erhalten wir

$$(3.50) \quad E(e'e) = \sigma^2\ Sp\ [I_n - (X'X)^{-1}X'X] = \sigma^2(Sp\ I_n - Sp\ I_k)$$

wegen (3.3).

Da somit I_n n und I_k k Hauptdiagonalelemente mit Wert 1 haben, erhalten wir aus (3.50)

$$(3.51) \quad E(e'e) = \sigma^2(n - k)$$

oder

$$(3.52) \quad \frac{E(e'e)}{n - k} = \sigma^2.$$

Mit (3.52) erkennen wir unmittelbar, daß für kleine Stichproben (kleine n) die Summe der Quadrate der Residuen eine verzerrte Schätzung der Summe der Quadrate der latenten Variablen ist.

Bilden wir darüberhinaus vergleichsweise den Erwartungswert der Summe der Quadrate der einzelnen latenten Variablen u_i (für i=1,...,n), so erhalten wir analog zu (3.46) und (3.48) mit (2.19)

$$(3.53) \quad E(u'u) = E[Sp(u\,u')] = Sp[E(u\,u')] = Sp(\sigma^2\ I_n)$$
$$= \sigma^2\ Sp\ (I_n) = \sigma^2\ .\ n$$

oder

$$(3.54) \quad \frac{E(u'u)}{n} = \sigma^2.$$

Demnach ist wegen (3.52) und (3.54)

$$(3.55) \qquad \frac{n}{n-k} E(e'e) = E(u'u).$$

Dann gilt

$$(3.56) \qquad E\left(\frac{n}{n-k} e'e\right) = E(u'u).$$

Also ergibt nur die mit dem Bereinigungsfaktor $\frac{n}{n-k}$ multiplizierte Streuung der Residuen der gewöhnlichen Methode der kleinsten Quadrate den gleichen Erwartungswert wie die Streuung der latenten Variablen unter den gegebenen Voraussetzungen.

Durch Vergleich von (3.56) mit (3.36) sehen wir, daß e'e eine verzerrte Schätzung von u'u ist und daß unter den gegebenen Voraussetzungen $\frac{n}{n-k}(e'e)$ eine erwartungstreue Schätzung von u'u ist.

In Skalarschreibweise bedeutet dies:
σ_u^2 ist zu schätzen mittels der Formel

$$(3.57) \qquad \hat{\sigma}_u^2 = \frac{\sum_{i=1}^{n} e_i^2}{n} \cdot \frac{n}{n-k} = \frac{\sum_{i=1}^{n} e_i^2}{n-k}.$$

Abschließend soll noch darauf hingewiesen werden, daß die Eigenschaften der gewöhnlichen Methode der kleinsten Quadrate von der tatsächlichen Erfüllung der Voraussetzungen abhängen. Dieser wesentliche Aspekt wird im zweiten Teil dieser Arbeit stärker herausgearbeitet.

4. Die verallgemeinerte Methode der kleinsten Quadrate[1)]

Die verallgemeinerte Methode der kleinsten Quadrate wurde von Alexander Craig Aitken entwickelt und in seinem Aufsatz "On Least Squares and Linear Combination of Observations" in den Proceedings of the Royal Society of Edinburgh, 55(1934 - 1935), auf den Seiten 42-48 dargelegt.

Die besondere Bedeutung dieser Methode besteht darin, daß über die Streuungs-Kovarianz-Matrix weniger restriktive Annahmen gemacht werden müssen. Diese Methode ist also z.B. für den Fall der Heteroskedastizität geeignet, d.h. wenn die Elemente auf der Hauptdiagonalen der Streuungs-Kovarianz-Matrix der latenten Variablen voneinander verschieden sind. Sie ist auch für den Fall geeignet, wenn die Kovarianzen von Null verschieden sind, also Autokorrelation vorliegt. Das hat zur Folge, daß für die Schätzung mit der Methode der kleinsten Quadrate nicht mehr die Voraussetzungen der Homoskedastizität und des Diagonalfalls (nur die Elemente auf der Hauptdiagonalen sind von Null verschieden) gegeben sein müssen. Wie bei den anderen Schätzmethoden für Gleichungssysteme muß jedoch angenommen werden, daß die Streuungs-Kovarianz-Matrix regulär ist, damit ihre Invertierbarkeit gewährleistet ist.

Wir gehen wieder von folgender reduzierte-Form-Matrizengleichung aus[2)]

---

1) Auch in diesem Kapitel werden aus schreibtechnischen Gründen Matrizen und Vektoren nicht durch andere Symbole gekennzeichnet. Dagegen werden die seltener vorkommenden Skalare besonders erwähnt.

2) Die Darstellung wurde angelehnt an Johnston: a.a.O., S. 179-183. Die Darstellung von Johnston ist ausführlicher als die Originaldarstellung von Aitken und verwendet [Fortsetzung S. 36.]

(2.16) $y = X\beta + u,$

wobei diese Vektoren und Matrizen wieder folgende Skalare enthalten

(2.17) $y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, $X = \begin{bmatrix} x_{11} & \cdots & x_{k1} \\ \vdots & & \vdots \\ x_{1n} & \cdots & x_{kn} \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$ und $u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$

Wir nehmen nun an

(2.18) $E(u) = 0,$

d.h. der Erwartungswert des Vektors der latenten Variablen sei der Nullvektor der Ordnung (n x 1).

Außerdem sei V die Streuungs-Kovarianz-Matrix der latenten Variablen

(4.1) $E(u\,u') = V,$

die regulär sein soll, d.h.

(4.2) $\det V \neq 0$, wobei Null hier ein Skalar ist.

V hat also nicht nur die Gestalt, die beim Diagonalfall mit Homoskedastizität zugrundegelegt wird:

---

zudem nicht die zum Teil irreführenden Symbole der Darstellung von Aitken. Obwohl in der Darstellung von Johnston nicht darauf hingewiesen wird, daß die verallgemeinerte Methode der kleinsten Quadrate auf Aitken zurückgeht, handelt es sich doch eindeutig um die Methode Aitkens. Vgl. hierzu Theil, H.: Economic Forecasts and Policy. 2. Aufl., Amsterdam 1961, S. 220 (1. Aufl: Amsterdam 1958).

$$(2.19)\quad V = \sigma^2 \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{bmatrix} = \sigma^2 I_n = E(u\,u')$$

V kann vielmehr jede beliebige Gestalt annehmen. Wenn wir jedoch eine bestimmte Gestalt von V auserwählen, so muß (4.1) auf jeden Fall Gültigkeit haben, d.h. V, die Streuungs-Kovarianz-Matrix der latenten Variablen, muß der Erwartungswert von u u' sein.

Wir suchen nun die Schätzung $\hat{b}$ des Parametervektors β und den hierzu erforderlichen linearen Regressor A, für die gilt[1)]

$$(4.3)\quad \hat{b} = A\,y,$$

wobei $\hat{b}$ und A folgende Skalare enthalten

$$(4.4)\quad \hat{b} = \begin{bmatrix} \hat{b}_1 \\ \vdots \\ \hat{b}_K \end{bmatrix} \quad \text{und} \quad A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{K1} & \cdots & A_{Kn} \end{bmatrix}.$$

---

1) Bei der gewöhnlichen Methode der kleinsten Quadrate lautete die Schätzung $\hat{\beta}$ des Parametervektors β

$$(2.29)\quad \hat{\beta} = (X'X)^{-1} X'y.$$

Daher gilt dort für den linearen Regressor A

$$(3.34)\quad A = (X'X)^{-1} X'$$

und somit auch

$$\hat{\beta} = A\,y \quad \text{(vgl. 3.35)}.$$

A soll also ein in den y-Werten linearer Regressor für b sein. Wenn A zudem ein unverzerrter linearer Regressor sein soll, so muß gelten

$$(4.5) \qquad E(A\,y) = E(\hat{b}) = \beta.$$

Durch Einsetzen von (2.16) erhalten wir

$$(4.6) \qquad \begin{aligned} E(A\,y) &= E[A(X\,\beta + u)] \\ &= E(A\,X\,\beta + A\,u) \\ &= A\,X\,\beta + A\,u \qquad \text{wegen (2.18).} \end{aligned}$$

(4.5) gilt also dann und nur dann, wenn[1)]

$$(4.7) \qquad A\,X = I_k.$$

(4.7) ist demnach eine hinreichende Bedingung für die Erwartungstreue von $\hat{b}$, d.h. daß die Schätzung $\hat{b}$ in (4.3) unverzerrt ist.

Wir wollen nun feststellen, ob auch für die verallgemeinerte Methode der kleinsten Quadrate die Summe der Varianzen der Schätzfehler minimal ist und wir bilden den Erwartungswert der Streuung von $\hat{b}$ um den "wahren" Vektor β, wobei wir (4.5) und (4.6) zugrundelegen.

$$(4.8) \qquad \begin{aligned} E[(\hat{b}-\beta)'(\hat{b}-\beta)] &= E[(A\,X\,\beta + A\,u - \beta)'(AX\beta + A\,u - \beta)] \\ &= E[(Au)'(Au)], \qquad \text{wegen (4.7)} \\ &= E\,(u'A'A\,u) \end{aligned}$$

Dies ist ein Skalar. Wir bilden die Spur[2)] dieses Skalars, die trivialerweise gleich dem Skalar selbst

---

1) I_k ist die Einheitsmatrix der Ordnung k.

2) Die Spur (Sp) einer quadratischen Matrix ist die Summe der Elemente auf der Hauptdiagonalen dieser Matrix. Vgl. Menges (1961): a.a.O., S. 224.

ist[1]. Es gilt also

(4.9) $E(u'A'a\ u) = Sp\ [E(u'A'A\ u)]$

Wir zerlegen nun $(u'A'A\ u)$ in $(u')\ (A'A\ u)$ und stellen fest[2], daß (u') von der Ordnung $(1 \times n)$ und $(A'A\ u)$ von der Ordnung $(n \times 1)$ sind. Dann gilt[3]

(4.10) $Sp[(u')(A'A\ u)] = Sp[(A'A\ u)(u')]$

oder, nach Wegfall der inneren Klammern,

(4.11) $Sp(u'A'A\ u) = Sp(A'A\ u\ u')$.

Ebenfalls gilt bei Berücksichtigung der Erwartungswerte

(4.12) $Sp[E(u'A'A\ u)] = Sp[E(A'A\ u\ u')]$.

Also erhalten wir für $E(u'A'A\ u)$

(4.13) $E(u'A'A\ u) = Sp[E(u'A'A\ u)] = Sp[E(A'A\ u\ u')]$

$= Sp(A'A\ V)$ wegen (4.1).

(4.8) ist also dann und nur dann die minimale Summe der Varianzen der Schätzfehler, wenn (4.13) ein Minimum erreicht, wenn also gilt

---

1) Ein Skalar kann als eine Matrix vom Rang 1×1 angesehen werden, die nur den Skalar als einziges Element enthält. Die Spur einer solchen degenerierten Matrix enthält also nur dieses Element, also den Skalar selbst.

2) Ein Spaltenvektor mit n Elementen kann als eine Matrix vom Rang $n \times 1$ angesehen werden, ein Zeilenvektor mit n Elementen entsprechend als eine Matrix vom Rang $1 \times n$.

3) Vgl. Zurmühl: a.a.O., S. 22.

(4.14) $\quad Sp(A'A\ V) \rightarrow min!$

A ist also so zu wählen, daß Sp(A'A V) minimiert wird unter der Restriktion $A\ X = I_k$.

Es ist also folgende Funktion Φ zu minimieren

(4.15) $\quad \Phi = Sp(A'A\ V) - Sp[M(A\ X - I_k)']$

wobei M die Matrix der Lagrangemultiplikatoren ist, die folgende Skalare enthält

(4.16) $\quad M = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1k} \\ \vdots & & \vdots \\ \mu_{k1} & \cdots & \mu_{kk} \end{bmatrix}$

Die partielle Ableitung von (4.15) nach A und Gleichsetzen mit Null (im Minimum ist die 1. Ableitung gleich Null) ergeben[1)]

(4.17) $\quad \frac{\partial \Phi}{\partial A} = 2\ kAV - kMX' = 0,$

wobei k der Skalar ist, der die Zahl der erklärenden Variablen angibt, und 0 die Nullmatrix der Ordnung (k × n) ist. Aus (4.17) erhalten wir

(4.18) $\quad 2\ AV = MX'.$

Durch Multiplikation von rechts mit V^{-1} X erhalten wir

---

1) Vgl. die Betrachtungen zu den Gleichungen (3.26) bis (3.30).

(4.19) $\quad 2\,AVV^{-1}\,X = MX'V^{-1}\,X$

und wegen $VV^{-1} = I_n$ und (4.7)

(4.20) $\quad 2\,I_k = MX'V^{-1}\,X.$

Durch Multiplikation von rechts mit $(X'V^{-1}X)^{-1}$ erhalten wir

(4.21) $\quad M = 2(X'V^{-1}X)^{-1}.$

Wir verwenden nun (4.18) und es ergibt sich

(4.22) $\quad 2\,AV = 2(X'V^{-1}X)^{-1}X'.$

Aus einer Multiplikation von rechts mit V^{-1} geht hervor

(4.23) $\quad A = (X'V^{-1}X)^{-1}X'V^{-1}.$

Wenn wir dies in (4.3) einsetzen, gelangen wir zu $\hat{b}$.

(4.24) $\quad \hat{b} = (X'V^{-1}X)^{-1}X'V^{-1}y.$

Dies ist also die Schätzung des Parametervektors, nämlich $\hat{b}$, mit der verallgemeinerten Methode der kleinsten Quadrate, die unter den angegebenen Bedingungen die minimale Summe der Varianzen der Schätzfehler im Vergleich zu allen anderen linearen erwartungstreuen Schätzungen aufweist.

Im nicht autokorrelierten Fall und zugleich bei Homoskedastizität[1] sind die verallgemeinerte Methode der klein-

1) Vgl. Johnston: a.a.O., S. 184, wo fälschlicherweise nur der nicht autokorrelierte Fall allein genannt wird.

sten Quadrate und die gewöhnliche Methode der kleinsten Quadrate identisch, da $V^{-1} = \frac{1}{\sigma^2} I_n$ und sich der Skalar $\frac{1}{\sigma^2}$ im Zähler und im Nenner von (4.23) wegkürzt und wir erhalten

$$(3.34) \qquad A = (X'X)^{-1}X'.$$

Wir suchen nun die Streuungs-Kovarianz-Matrix der Parameterschätzung $\hat{b}$. Aus (4.8) können wir herleiten, daß

$$(4.25) \quad E[(\hat{b}-\beta)(\hat{b}-\beta)'] = E[(Au)(Au')] = A[E(u\,u')]A' = AVA' .$$

Dann gilt wegen (4.25) und (4.23)

$$(4.26) \quad E[(\hat{b}-\beta)(\hat{b}-\beta)'] = (X'V^{-1}X)^{-1}X'V^{-1}VV^{-1}X(X'V^{-1}X)^{-1}.$$

Somit erhalten wir wegen $VV^{-1} = I_n$ und $(X'V^{-1}X)^{-1}X'V^{-1}X = I_k$

$$(4.27) \qquad E[(\hat{b} - \beta)(\hat{b} - \beta)'] = (X'V^{-1}X)^{-1}$$

Dies ist der Ausdruck für die Streuungs-Kovarianz-Matrix der Parametervektorschätzung $\hat{b}$.

Wir sehen auch hier, daß bei Homoskedastizität und zugleich Diagonalfall die Streuungs-Kovarianz-Matrix der Parametervektorschätzung nach der verallgemeinerten Methode der kleinsten Quadrate mit der nach der gewöhnlichen Methode der kleinsten Quadrate[1)] übereinstimmt, da dann $V^{-1} = \frac{1}{\sigma^2} I_n$ und $\frac{1}{\frac{1}{\sigma^2}} = \sigma^2$ ist.

---

1) Vgl. Gleichung (3.12).

5. Die zweistufige Methode der kleinsten Quadrate[1)]

Die zweistufige Methode der kleinsten Quadrate wurde von Henri Theil[2)] entwickelt. Sie ist eine Methode zur Schätzung der Parameter einer einzelnen Gleichung eines größeren Gleichungssystems. Diese Gleichung kann außer den vorherbestimmten Variablen und der latenten Variablen mehrere unverzögerte endogene Variablen enthalten.

Von diesen gemeinsam abhängigen Variablen der Gleichung wird nun eine Variable ausgewählt und die Gleichung nach dieser Variablen aufgelöst. Diese ausgewählte endogene Variable hängt also ab von den anderen unverzögerten endogenen Variablen, den vorherbestimmten Variablen und der latenten Variablen. Unsere Gleichung, es sei z.B. die erste des größeren, sie enthaltenden Systems lautet dann in Matrizenschreibweise[3)]

$$(5.1) \qquad y_1 = -Y_2\beta_{12}' - X_*\gamma_{1*}' + u_1$$

wobei diese Matrizen und Vektoren folgende Skalare enthalten:

$$(5.2) \quad y_1 = \begin{bmatrix} y_{11} \\ \cdot \\ \cdot \\ \cdot \\ y_{1n} \end{bmatrix}, Y_2 = \begin{bmatrix} y_{21} & \cdots & y_G{}_{\Delta_1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ y_{2n} & \cdots & y_G{}_{\Delta_n} \end{bmatrix}, \beta_{12}' = \begin{bmatrix} \beta_{12} \\ \cdot \\ \cdot \\ \cdot \\ \beta_{1G}{}_{\Delta} \end{bmatrix},$$

1) Die Darstellung entspricht der von Johnston: a.a.O., S. 258-260.

2) Theil (1958): a.a.O., S. 223-229
und Theil, H.: The aggregation implications of identifiable structural macrorelations. Econometrica, 27(1959), S. 14-29, s. bes. S. 16-18.

3) Aus schreibtechnischen Gründen werden Skalare besonders erwähnt, jedoch Matrizen und Vektoren in normalen Symbolen geschrieben.

$$X_* = \begin{bmatrix} x_{11} & \cdots & x_{K*1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_{1n} & \cdots & x_{K*n} \end{bmatrix}, \quad \gamma'_{1*} = \begin{bmatrix} \gamma_{11} \\ \cdot \\ \cdot \\ \gamma_{1K*} \end{bmatrix} \text{ und } u_1 = \begin{bmatrix} u_{11} \\ \cdot \\ \cdot \\ u_{1n} \end{bmatrix}$$

Hierbei ist Y_2 die Matrix der Beobachtungswerte derjenigen gemeinsam abhängigen Variablen, die in dieser Gleichung vorkommen, nach denen (5.2) aber nicht aufgelöst wurde. β'_{12} ist der zugehörige Parametervektor. X_* enthält die Beobachtungswerte derjenigen vorherbestimmten Variablen, die in dieser Gleichung enthalten sind; der Vektor γ'_{1*} enthält deren zugehörige Parameter; u_1 ist der Vektor der latenten Variablen dieser ersten Strukturgleichung. Enthält das Gleichungssystem G gemeinsam abhängige und K vorherbestimmte Variablen, so sind demnach $G - G^{\Delta} = G^{\Delta\Delta}$ gemeinsam abhängige und $K - K^* = K^{**}$ vorherbestimmte Variablen aus dieser Gleichung ausgeschlossen, wobei G, G^{Δ}, $G^{\Delta\Delta}$, K, K*, K** Skalare sind.

1. Wir nehmen nun an, daß uns für n Perioden die Werte der gemeinsam abhängigen und vorherbestimmten Variablen bekannt sind.

2. Unsere Gleichung sei identifizierbar, d.h.

(5.3) $\quad K^{**} \geq G^{\Delta} - 1$

3. Außerdem nehmen wir an, daß die n nicht beobachtbaren latenten Variablen $u_{11}, \ldots, u_{1n}$ unabhängig sind von den $n \times K^*$ Werten der vorherbestimmten Variablen. Demnach können die latenten Variablen nur noch von den gemeinsam abhängigen Variablen abhängig sein. Die zweistufige Methode der kleinsten Quadrate versucht gerade , den Einfluß

der nicht ausgewählten gemeinsam abhängigen Variablen auf die latenten Variablen zu beseitigen.

Schließlich gelte nach wie vor [vgl. (2.18) und (2.19)]

(5.4) $E(u_1) = 0$, wobei 0 für den Nullspaltenvektor der Ordnung (n x 1) steht, und

(5.5) $E(u_1 u_1') = \sigma^2 I_n$.

Die zweistufige Methode der kleinsten Quadrate schätzt nun in der ersten Stufe Y_2 in Gleichung (5.1) als Funktion der Matrix aller vorherbestimmten Variablen X, wobei die Matrix X folgende Skalare enthält

(5.6) $$X = \begin{bmatrix} x_{11} & \cdots & x_{K1} \\ \vdots & & \vdots \\ x_{1n} & \cdots & x_{Kn} \end{bmatrix}$$

Wir nehmen also an, Y_2 sei eine lineare ökonometrische Funktion von X. Dann gelte folgendes System von $G^{\Delta} - 1$ reduzierte-Form-Gleichungen

(5.7) $$Y_2 = X \alpha' + v ,$$

wobei die Matrizen α' und v folgende Skalare enthalten

(5.7a) $$\alpha' = \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{G^{\Delta}1} \\ \vdots & & \vdots \\ \alpha_{1K} & \cdots & \alpha_{G^{\Delta}K} \end{bmatrix} \qquad v = \begin{bmatrix} v_{21} & \cdots & v_{G^{\Delta}1} \\ \vdots & & \vdots \\ v_{2n} & \cdots & v_{G^{\Delta}n} \end{bmatrix}$$

Spätestens jetzt müssen wir die Annahme treffen, daß auch die latenten Variablen der reduzierte-Form-Gleichungen von den vorherbestimmten Variablen stochastisch unabhängig sind, d.h. wir müssen fordern, daß

$$(5.8) \qquad E(X'v) = 0,$$

wobei 0 für die Nullmatrix der Ordnung $(K \times G^{\Delta})$ steht. Außerdem muß vorausgesetzt werden, daß die reduzierte Form überhaupt existiert. Schließlich muß auch für v gelten

$$(5.9) \qquad E(v) = 0,$$

wobei 0 für die Nullmatrix der Ordnung $[n \times (G^{\Delta}-1)]$ steht, sowie

$$(5.10) \qquad E(v\,v') = \sigma_v^2\, I_n,$$

wobei σ_v^2 ein Skalar ist.

Wir minimieren nun die Quadrate der Abweichungen der reduzierte-Form-Schätzgleichungen (5.11)

$$(5.11) \qquad Y_2 - \hat{Y}_2 = Y_2 - X\,\hat{\alpha}' = w$$

wobei $\hat{Y}_2$, $\hat{\alpha}$ und w folgende Skalare enthalten

(5.12)

$$\hat{Y}_2 = \begin{bmatrix} \hat{y}_{21} & \cdots & \hat{y}_{G\Delta_1} \\ \vdots & & \vdots \\ \hat{y}_{2n} & \cdots & \hat{y}_{G\Delta_n} \end{bmatrix}, \quad \hat{\alpha}' = \begin{bmatrix} \hat{\alpha}_{21} & \cdots & \hat{\alpha}_{G\Delta_1} \\ \vdots & & \vdots \\ \hat{\alpha}_{2K} & \cdots & \hat{\alpha}_{G\Delta_K} \end{bmatrix} \text{ und } w = \begin{bmatrix} w_{21} & \cdots & w_{G\Delta_1} \\ \vdots & & \vdots \\ w_{2n} & \cdots & w_{G\Delta_n} \end{bmatrix}$$

Die Minimierung der Quadrate der Abweichungen der reduzierte-Form-Gleichungen erfolgt am übersichtlichsten

gleichungsweise. Wir minimieren also separat
$w_2' w_2$, ..., $w_{G\Delta}' w_{G\Delta}$, wobei w_2, ..., $w_{G\Delta}$ folgende Skalare enthalten

$$(5.13) \quad w_2 = \begin{bmatrix} w_{21} \\ \cdot \\ \cdot \\ \cdot \\ w_{2n} \end{bmatrix}, \ \ldots, \ w_{G\Delta} = \begin{bmatrix} w_{G\Delta_1} \\ \cdot \\ \cdot \\ \cdot \\ w_{G\Delta_n} \end{bmatrix},$$

und zwar jeweils nur bezüglich des zugehörigen Spaltenvektors aus $\hat{\alpha}$, wenn $\hat{\alpha}$ in folgende Spaltenvektoren zerfällt: $\hat{\alpha}_2$, ..., $\hat{\alpha}_{G\Delta}$, welche folgende Skalare enthalten

$$(5.14) \quad \hat{\alpha}_2' = \begin{bmatrix} \hat{\alpha}_{21} \\ \cdot \\ \cdot \\ \cdot \\ \hat{\alpha}_{2K} \end{bmatrix}, \ \ldots, \ \hat{\alpha}_{G\Delta}' = \begin{bmatrix} \hat{\alpha}_{G\Delta_1} \\ \cdot \\ \cdot \\ \cdot \\ \hat{\alpha}_{G\Delta_K} \end{bmatrix}.$$

Durch Zusammenfassung der reduzierte-Form-Schätzvektoren $\hat{\alpha}_2'$, ..., $\hat{\alpha}_{G\Delta}'$ nach in üblicher Weise[1] erfolgter Minimierung zur reduzierte-Form-Schätzmatrix $\hat{\alpha}$, erhalten wir schließlich

$$(5.15) \quad \hat{\alpha}' = (X'X)^{-1} X'Y_2$$

und damit die Schätzung von Y_2, nämlich $\hat{Y}_2$ unter Zuhilfenahme von (5.11)

$$(5.16) \quad \hat{Y}_2 = X(X'X)^{-1} X'Y_2.$$

Dann gilt für w

1) Vgl. die Betrachtungen von Gleichung (2.26) bis Gleichung (2.29).

(5.17) $$w = Y_2 - X(X'X)^{-1} X'Y_2.$$

Wir können dies umformen zu

(5.18) $$Y_2 - w = X(X'X)^{-1} X'Y_2 = \hat{Y}_2.$$

Das Ergebnis der ersten Stufe ist also die Gleichung

(5.19) $$y_1 = -\hat{Y}_2 \beta'_{12} - X_* \gamma'_{1*} + w_1.$$

Was sich hinter w_1 verbirgt, erfahren wir, wenn wir in Gleichung (5.1) gleichzeitig auf der rechten Seite $w \beta'_{12}$ addieren und subtrahieren, wodurch sich der Wert nicht ändert:

(5.20) $$y_1 = -(Y_2 - w)\beta'_{12} - X_* \gamma'_{1*} + (u_1 - w \beta'_{12})$$

Hieraus ergibt sich wegen (5.18)

(5.21) $$y_1 = -\hat{Y}_2 \beta'_{12} - X_* \gamma'_{1*} + (u_1 - w \beta'_{12}).$$

Durch Vergleich mit (5.19) erkennen wir, daß

(5.22) $$w_1 = u_1 - w \beta'_{12}.$$

Die latente Variable von Gleichung (5.19) ergibt sich also aus der Differenz zwischen dem Vektor der latenten Variablen von (5.1) und dem Produkt der Matrix der reduzierte-Form-Abweichungen, w, mit dem Parametervektor β'_{12}.

Zur Vorbereitung der zweiten Stufe verkürzen wir (5.20) zu

(5.23) $$y_1 = -A \delta' + w_1$$

wobei die Matrix A aus den folgenden beiden Untermatrizen besteht

$$(5.24) \qquad A = [(Y_2 - w) \quad X_*]$$

und der Vektor δ' aus folgenden beiden Untervektoren besteht

$$(5.24) \qquad \delta' = \begin{bmatrix} \beta'_{12} \\ \gamma'_{1*} \end{bmatrix}.$$

Die zweite Stufe besteht nun darin, den Parametervektor δ' in (5.25) zu schätzen mit der gewöhnlichen Methode der kleinsten Quadrate. Wir erhalten

$$(5.26) \qquad \hat{\delta}' = -(A'A)^{-1}A'y_1.$$

Durch Einsetzen von (5.24) und (5.25) in (5.26) erhalten wir schließlich

$$(5.27) \qquad \begin{bmatrix} \hat{\beta}'_{12} \\ \hat{\gamma}'_{1*} \end{bmatrix} = - \begin{bmatrix} (Y_2-w)'(Y_2-w) & (Y_2-w)'X_* \\ X_*'(Y_2-w) & X_*'X_* \end{bmatrix}^{-1} \begin{bmatrix} Y_2'-w' \\ X_*' \end{bmatrix} y_1.$$

Nun ist aber wegen (5.17)

$$(5.28) \qquad X'w = X'Y_2 - X'X(X\,X)^{-1}\,X'Y_2 = 0$$

wobei 0 für die Nullmatrix der Ordnung $[K \times (G^{\Delta} - 1)]$ steht, und $X_*'w$, eine Untermatrix dieser Nullmatrix, muß daher ebenfalls eine Nullmatrix sein. Entsprechend läßt sich dies für $w'X_*$ zeigen.

Außerdem erhalten wir für

$$(5.29)\qquad (Y_2-w)'(Y_2-w) = Y_2' Y_2 - Y_2' w - w' Y_2 + w'w$$
$$= Y_2' Y_2 - w'w - w'w + w'w$$
$$= Y_2' Y_2 - w'w,$$

denn es gelten wegen (5.11)

$$(5.30)\qquad Y_2' w = (\hat{Y}_2 + w)'w = \hat{Y}_2' w + w'w = w'w$$

und

$$(5.31)\qquad w'Y_2 = w'(\hat{Y}_2+w) = w'\hat{Y}_2 + w'w = w'w,$$

weil

$$(5.32)\qquad \hat{Y}_2' w = 0_G{}^{\Delta}{}_{-1} = w'\hat{Y}_2,$$

wobei $0_G{}^{\Delta}{}_{-1}$ die Nullmatrix der Ordnung $(G^{\Delta} -1)$ ist. (5.32) besagt nichts anderes, als daß die Summe der Produkte von Kleinstquadratschätzungen mit ihren jeweiligen Residuen immer gleich Null ist. Wir beweisen die Gültigkeit von (5.32) mit Hilfe von (5.11) und (5.16):

$$(5.33)\qquad \hat{Y}_2' w = \hat{Y}_2' (Y_2 - \hat{Y}_2)$$
$$= Y_2' X(X'X)^{-1} X'[Y_2 - X(X'X)^{-1} X'Y_2]$$
$$= Y_2' X(X'X)^{-1} X'Y_2 - Y_2' X(X'X)^{-1} X'Y_2$$
$$= 0_G{}^{\Delta}{}_{-1}$$

Mit Hilfe der Vereinfachungen von (5.28) bis (5.29) können wir jetzt für (5.27) schreiben

$$(5.34)\qquad \begin{bmatrix} \hat{\beta}_{12}^{*} \\ \hat{\gamma}_{1*}^{*} \end{bmatrix} = - \begin{bmatrix} Y_2'Y_2 - w'w & Y_2' X_* \\ X_*' Y_2 & X_*' X_* \end{bmatrix}^{-1} \begin{bmatrix} Y_2' - w' \\ X_*' \end{bmatrix} y_1 .$$

Dies ist die Schätzung nach der zweistufigen Methode der kleinsten Quadrate.

6. Die dreistufige Methode der kleinsten Quadrate

Die dreistufige Methode der kleinsten Quadrate wurde von Arnold Zellner und Henri Theil[1] entwickelt. Diese Methode ist zur Simultanschätzung eines Gleichungssystems konzipiert worden.

Wir gehen[2] zunächst von der ersten Gleichung eines solchen Gleichungssystems aus und betrachten erst später das Gleichungssystem in seiner Gesamtheit.

Die erste Gleichung des Gleichungssystems habe folgendes Aussehen[3]

$$(6.1) \qquad y_1 = Y_2 \beta_{12}' + X_* \gamma_{1*}' + u_1,$$

wobei die in (6.1) enthaltenen Vektoren und Matrizen in (5.2) definiert sind. In der Tat unterscheidet sich (6.1) von (5.1) lediglich hinsichtlich des Vorzeichens von $Y_2 \beta_{12}'$ und von $X_* \gamma_{1*}'$.

Die dreistufige Methode der kleinsten Quadrate basiert nun auf folgenden Voraussetzungen:

1) Wir setzen voraus, daß wir für n Perioden Beobachtungswerte aller endogenen und vorherbestimmten Variablen des Gleichungssystems zur Verfügung haben.

---

1) Zellner, A. u. Theil, H.: Three-stage least squares: Simultaneous estimation of simultaneous equations. Econometrica, 30(1962), S. 54-78.

2) Vgl. zur Darstellung: Zellner u. Theil: a.a.O., S. 54-58 sowie Johnston: a.a.O., S. 266-268.

3) Die Symbole stellen Vektoren und Matrizen dar, wenn sie nicht besonders als Skalare gekennzeichnet oder aus dem Zusammenhang als Skalare erkennbar sind.

2) Jede Gleichung des Systems sei identifizierbar, d.h. für jede Gleichung des Gleichungssystems gelte

(6.2) $K^{**} \geqq G^{\Delta} - 1.$

3) Die Parametermatrix der endogenen Variablen des Gleichungssystems sei regulär, so daß die reduzierte Form über das Invertieren dieser Matrix berechnet werden kann.

4) Der Erwartungswert aller latenten Variablen des Strukturgleichungssystems sei Null.

5) Die latenten Variablen jeder Strukturgleichung seien nicht autokorreliert und homoskedastisch in dem Sinn, daß ihre Streuung zeitlich konstant ist. Dagegen müssen die Streuungen der latenten Variablen verschiedener Strukturgleichungen nicht miteinander übereinstimmen. Die Kovarianzen der latenten Variablen verschiedener Gleichungen zum gleichen Zeitpunkt seien ebenfalls konstant über die Zeit aber nicht identisch über alle Strukturgleichungen.

   Die sich aus diesen Annahmen ergebende Gestalt der Streuungs-Kovarianz-Matrix der latenten Variablen aller Strukturgleichungen sehen wir für die reduzierte Form in (6.20) mit (6.21).

6) Die Varianz-Kovarianz-Matrix der latenten Variablen aller Strukturgleichungen sei regulär.

7) Die vorherbestimmten Variablen seien stochastisch unabhängig von den latenten Variablen.

Durch Multiplikation beider Seiten von Gleichung (6.1) mit X' von links erhalten wir

$$(6.3) \qquad X'y_1 = X'Y_2\beta_{12}' + X'X_*\gamma_{1*}' + X'u_1,$$

wobei die Matrix X wie in (5.6) definiert ist.
(6.3) können wir vereinfachen zu

$$(6.4) \qquad X'y_1 = X'Z_1\delta_1' + X'u_1,$$

wobei

$$(6.5) \qquad Z_1 = [Y_2 \; X_*] \text{ und } \delta_1' = \begin{bmatrix} \beta_{12}' \\ \gamma_{1*}' \end{bmatrix}.$$

Wir vereinfachen die Schreibweise von (6.4) weiter zu

$$(6.6) \qquad A_1 = B_1 \delta_1' + v_1$$

wobei

$$(6.7) \qquad A_1 = X'y_1, \quad B_1 = X'Z_1 \text{ und } v_1 = X'u_1.$$

Wir stellen fest, daß auch der Erwartungswert des in (6.7) definierten Vektors der latenten Variablen, v_1, von (6.6) Null ist wegen Voraussetzung 7.

Wenn (6.6) <u>exakt identifiziert</u> ist, können wir die Parameterschätzung $\hat{\delta}_1'$ mit der indirekten Methode der kleinsten Quadrate lediglich durch die algebraische Umformung

$$(6.8) \qquad \begin{aligned} \hat{\delta}_1' &= B_1^{-1} A_1 \\ &= (X' Z_1)^{-1} X'y_1 \end{aligned}$$

berechnen.

Da jedoch im allgemeinen <u>Überidentifikation</u> vorliegt, wenden wir die verallgemeinerte Methode der kleinsten Quadrate an. Diese verwendet die Streuungs-Kovarianz-Matrix der latenten Variablen von (6.6)

$$(6.9) \qquad V_{11} = E(v_1 v_1') = E(X' u_1 u_1' X) = \sigma_{11}(X'X)$$

wegen (6.7) und Voraussetzung 5, wonach σ_{11} ein Skalar ist, der die zeitlich konstante Streuung der latenten Variablen von (6.1) repräsentiert, während die Kovarianzelemente gleich Null sind.

Da die Schätzformel nach der verallgemeinerten Methode der kleinsten Quadrate lautet[1)]

$$(6.10) \qquad \hat{\delta}_1' = (B_1' V_{11}^{-1} B_1)^{-1} B_1' V_{11}^{-1} A_1$$

erhalten wir durch Einsetzen von (6.7) und (6.9) in (6.10)

$$(6.11) \qquad \hat{\delta}_1' = [Z_1' X \sigma^{11}(X'X)^{-1} X' Z_1]^{-1} Z_1' X \sigma^{11}(X'X)^{-1} X' y_1,$$

wobei $\sigma^{11} = \frac{1}{\sigma_{11}}$.

Da σ^{11} ebenfalls ein Skalar ist, können wir σ^{11} vorziehen bei der Multiplikation und wir erhalten, nachdem sich $\frac{1}{\sigma^{11}}$ gegen σ^{11} wegkürzt,

$$(6.12) \qquad \hat{\delta}_1' = [Z_1' X(X'X)^{-1} X' Z_1]^{-1} Z_1' X(X'X)^{-1} X' y_1.$$

Dies ist die Schätzung nach der zweistufigen Methode der kleinsten Quadrate. Dies sei kurz gezeigt:

Das aufgrund von X mit der gewöhnlichen Methode der kleinsten Quadrate geschätzte Z_1 lautet in der ersten Stufe[2)]

---

1) Vgl. Gleichung (4.24).
2) Vgl. (2.30).

$$(6.13) \qquad \hat{Z}_1 = X(X'X)^{-1} X'Z_1 .$$

In der zweiten Stufe erhalten wir die Parameterschätzung $\hat{\delta}_1'$ durch[1)]

$$(6.14) \qquad \hat{\delta}_1' = (\hat{Z}_1' \hat{Z}_1)^{-1} \hat{Z}_1' y_1 .$$

Durch Einsetzen von (6.13) in (6.14) erhalten wir

$$(6.15) \qquad \hat{\delta}_1' = [Z_1' X(X'X)^{-1} X'X(X'X)^{-1} X'Z_1]^{-1} Z_1' X(X'X)^{-1}X'y_1 ,$$

woraus wir wegen $(X'X)^{-1} X'X = I_k$ nun (6.12) erhalten.

Zur dritten Stufe der dreistufigen Methode der kleinsten Quadrate gelangen wir, wenn wir nicht nur die erste Gleichung sondern auch alle übrigen Gleichungen des Gleichungssystems in die Form von Gleichung (6.4) bringen. Dann sieht unser neues Gleichungssystem folgendermaßen aus

$$(6.16) \qquad \begin{bmatrix} X'y_1 \\ X'y_2 \\ \vdots \\ X'y_G \end{bmatrix} = \begin{bmatrix} X'Z_1 & 0 & \cdots & 0 \\ 0 & X'Z_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & X'Z_G \end{bmatrix} \begin{bmatrix} \delta_1' \\ \delta_2' \\ \vdots \\ \delta_G' \end{bmatrix} + \begin{bmatrix} X'u_1 \\ X'u_2 \\ \vdots \\ X'u_G \end{bmatrix}$$

oder, verkürzt,

$$(6.17) \qquad A = B\,\delta' + v$$

mit

$$(6.18) \quad A = \begin{bmatrix} X'y_1 \\ X'y_2 \\ \vdots \\ X'y_G \end{bmatrix}, \quad B = \begin{bmatrix} X'Z_1 & 0 & \cdots & 0 \\ 0 & X'Z_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & X'Z_G \end{bmatrix}, \quad \delta' = \begin{bmatrix} \delta_1' \\ \delta_2' \\ \vdots \\ \delta_G' \end{bmatrix} \text{ und } v = \begin{bmatrix} X'u_1 \\ X'u_2 \\ \vdots \\ X'u_G \end{bmatrix} .$$

1) Vgl. (2.29), wo auch β ein Spaltenvektor ist.

Diesmal lautet die Schätzung von δ mit Hilfe der verallgemeinerten Methode der kleinsten Quadrate

$$(6.19) \qquad \hat{\delta}' = [B' V^{-1} B]^{-1} B' V^{-1} A,$$

wobei V folgende Gestalt hat

$$(6.20) \quad V = E(v\,v') = \begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1G} \\ V_{21} & V_{22} & \cdots & V_{2G} \\ \vdots & \vdots & & \vdots \\ V_{G1} & V_{G2} & \cdots & V_{GG} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11}X'X & \sigma_{12}X'X & \cdots & \sigma_{1G}X'X \\ \sigma_{21}X'X & \sigma_{22}X'X & \cdots & \sigma_{2G}X'X \\ \vdots & \vdots & & \vdots \\ \sigma_{G1}X'X & \sigma_{G2}X'X & \cdots & \sigma_{GG}X'X \end{bmatrix},$$

wobei sich herleiten läßt

$$(6.21) \quad \sigma_{ij}X'X = \begin{bmatrix} \sigma_{ij} & 0 & \cdots & 0 \\ 0 & \sigma_{ij} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_{ij} \end{bmatrix} \cdot X'X$$

$$= \sigma_{ij} I_K \cdot X'X = X'(\sigma_{ij} I_n)X \text{ für } i,j=1,\ldots,G$$

wobei wieder σ_{ij} ein Skalar ist.

Einsetzen von (6.18) und (6.20) in (6.19) ergibt

$$(6.22)\quad \hat{\delta}' = \begin{bmatrix} Z_1' X \frac{1}{\sigma_{11}} (X'X)^{-1}X'Z_1 & \dots & Z_1' X \frac{1}{\sigma_{1G}} (X'X)^{-1}X'Z_G \\ \vdots & & \vdots \\ Z_G' X \frac{1}{\sigma_{G1}} (X'X)^{-1}X'Z_1 & \dots & Z_G' X \frac{1}{\sigma_{GG}} (X'X)^{-1}X'Z_G \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} Z_1' X \frac{1}{\sigma_{11}} (X'X)^{-1} & \dots & Z_1' X \frac{1}{\sigma_{1G}} (X'X)^{-1} \\ \vdots & & \vdots \\ Z_G' X \frac{1}{\sigma_{G1}} (X'X)^{-1} & \dots & Z' X \frac{1}{\sigma_{GG}} (X'X)^{-1} \end{bmatrix} \begin{bmatrix} X'y_1 \\ \vdots \\ X'y_G \end{bmatrix}$$

(6.22) ist das Ergebnis der Schätzung des Parameterspaltenvektors δ' durch $\hat{\delta}'$ nach der dreistufigen Methode der kleinsten Quadrate.

Die dreistufige Methode der kleinsten Quadrate stimmt mit der zweistufigen Methode der kleinsten Quadrate überein[1], wenn die Streuungs-Kovarianz-Matrix der latenten Variablen der Gleichung (6.17), welche in (6.20) definiert ist, nur V_{ii}-Elemente (i=1,...,G) enthält, d.h. wenn die Kovarianz-Matrizen V_{ij} für $i \neq j$ (i,j=1,...,G) Nullmatrizen der Ordnung k sind. Dann lassen sich die einzelnen Gleichungen vom Typ (6.12) sofort aus (6.22) herleiten, da dann auch die σ_{ij} für $i \neq j$ (i=1,...,G) dem Skalar Null gleich sind.

Da die in (6.20) enthaltenen Skalare σ_{ij} für i,j=1,...,G im allgemeinen nicht a priori spezifiziert sind, schlagen Zellner und Theil[2] vor, sie durch die Streuungen und Kovarianzen der nach dem zweiten Schritt erhaltenen Abweichungen[3] zu schätzen.

---

1) Vgl. Zellner u. Theil: a.a.O., S. 58.

2) Zellner u. Theil: a.a.O., S. 58.

3) Für Abweichungen ist auch der Ausdruck Residuen gebräuchlich.

7. Die indirekte Methode der kleinsten Quadrate[1)]

Die indirekte Methode der kleinsten Quadrate ist eine reduzierte-Form-Schätzmethode. Die reduzierte-Form-Parameter werden hierbei nach der gewöhnlichen Methode der kleinsten Quadrate geschätzt. Die indirekte Methode der kleinsten Quadrate kann zur Schätzung der Parameter einer Strukturgleichung verwandt werden, wenn folgende Bedingungen erfüllt sind:

1) Aus der Strukturgleichung muß die reduzierte Form berechnet werden können. Gehört die Strukturgleichung einem vollständigen Gleichungssystem an, so muß die Parametermatrix der endogenen Variablen aller Strukturgleichungen regulär sein, damit durch Inversion dieser Parametermatrix die reduzierte Form gewonnen werden kann.

2) Die reduzierte-Form-Parameter müssen geschätzt werden können. Dies wird, abgesehen von der Erfüllung fundamentaler Voraussetzungen wie etwa dem Vorhandensein genügend zahlreicher Beobachtungen, schon dann möglich sein, wenn die reduzierte-Form-Gleichungen identifizierbar sind.[2)]

3) Eine rechnerische Übertragung der reduzierte-Form-Parameterschätzungen in die Strukturgleichung muß gewährleistet sein. Aus den reduzierte-Form-Parameterschätzungen

---

1) Vgl. Koopmans, T.C. u. Hood, W.C.: The estimation of simultaneous linear economic relationships. In: Studies in Econometric Method. Hrsg.: W.C. Hood u. T.C. Koopmans. New York 1953, S. 112-199, s.bes. S. 142-143;
Menges (1961): a.a.O., S. 98-99;
Johnston: a.a.O., S. 253.

2) Daß jedes Modell in reduzierter Form (mit normal verteilten latenten Variablen) identifizierbar ist, wird in Menges (1961): a.a.O., S. 78-79 bewiesen.

müssen also die Strukturgleichungsparameterschätzungen berechnet werden können. Die Übertragung der reduzierte-Form-Parameterschätzungen auf die betrachtete Strukturgleichung ist jedoch dann und nur dann eindeutig durchzuführen, wenn die betrachtete Sturkturgleichung exakt identifizierbar ist.

Die dritte Bedingung stellt die wichtigste Einschränkung der Anwendbarkeit der indirekten Methode der kleinsten Quadrate dar. Da in der ökonometrischen Praxis in aller Regel die Strukturgleichungen überidentifiziert sind, kann die indirekte Methode der kleinsten Quadrate nicht angewandt werden, weil die Übertragung der reduzierte-Form-Parameterschätzungen auf die betrachtete Strukturgleichung nicht möglich ist.

Das Vorgehen der indirekten Methode der kleinsten Quadrate wird im Rahmen der Darstellung der Methode der größten Dichte bei beschränkter Information nach Anderson und Rubin[1] für den Unterfall exakter Identifikation[2] veranschaulicht. Gerade die letztere Methode zeichnet sich gegenüber der indirekten Methode der kleinsten Quadrate aus durch die Berücksichtigung überzähliger Informationen bei Überidentifikation und gelangt dann zu besseren Schätzergebnissen.

---

1) Vgl. Kapitel 9.

2) Bei exakter Identifikation führen beide Methoden zu gleichen Schätzergebnissen.

8. Die Methode der größten Dichte bei vollständiger Information

Die Methode der größten Dichte[1] wurde 1912 von Ronald Aylmer Fisher[2] entwickelt[3]. Die folgende Darstellung entspricht der von Koopmans, Rubin und Leipnik[4].
Wir gehen aus von dem Gleichungssystem[5]

$$(8.1) \qquad B\, y_t' + \Gamma\, z_t' = u_t' \qquad (t=1,\ldots,T)\ ,$$

wobei diese Matrizen und Vektoren folgende Skalare enthalten

$$(8.2) \quad B = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1G} \\ \vdots & & \vdots \\ \beta_{G1} & \cdots & \beta_{GG} \end{bmatrix},\ y_t' = \begin{bmatrix} y_{1t} \\ \vdots \\ y_{Gt} \end{bmatrix},\ \Gamma = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1K} \\ \vdots & & \vdots \\ \gamma_{G1} & \cdots & \gamma_{GK} \end{bmatrix},$$

$$z_t' = \begin{bmatrix} z_{1t} \\ \vdots \\ z_{Kt} \end{bmatrix} \quad \text{und} \quad u_t' = \begin{bmatrix} u_{1t} \\ \vdots \\ u_{Gt} \end{bmatrix}.$$

---

1) Die deutsche Bezeichnung "Methode der größten Dichte" für method of maximum likelihood wurde von G. Menges [(1961), a.a.O., S. 99-100] geprägt.

2) Fisher, R.A.: On an absolute criterion for fitting frequency curves. Messenger of Mathematics, 41(1912), S. 155 ff.

3) Vgl. Cramér, H.: Mathematical Methods of Statistics. Princeton 1946, S. 498.

4) Koopmans, T.C., Rubin, H. u. Leipnik, R.B.: Measuring the equation systems of dynamic economics. In: Statistical Inference in Dynamic Economic Models. Hrsg.: Tjalling C. Koopmans. New York-London 1950, S. 53-237, s. bes. S. 69-78 u. S. 110-120.
Vgl. auch Menges (1961): a.a.O., S. 99-104. Auch in diesem Kapitel werden nur Skalare besonders betont.

5) Koopmans, Rubin u. Leipnik, S. 70-72.

Wir können (8.1) weiter vereinfachen zu

$$(8.3) \qquad A\, x_t' = u_t' \qquad (t=1,\ldots,T),$$

wobei

$$(8.4) \qquad A = [B \quad \Gamma] \text{ und } x_t' = \begin{bmatrix} y_t' \\ z_t' \end{bmatrix}.$$

Die Unterscheidung zwischen gemeinsam abhängigen und vorherbestimmten Variablen kann bei diesem Verfahren vernachlässigt werden[1].

Wir nehmen nun an, daß die latenten Variabeln von den vorherbestimmten Variablen unabhängig sind. Aus dieser Annahme folgt[2]

$$(8.5) \qquad p(u_t' \mid z_t') = p(u_t') \qquad (t=1,\ldots,T).$$

Die Dichtefunktion der latenten Variablen wird somit durch die vorherbestimmten Variablen nicht beeinflußt.

Diese Annahme gestattet daher eine eindeutige Abbildung von den Elementen des Vektors u_t' auf die Elemente des Vektors y_t'. Die Funktion dieser Abbildung erhalten wir durch partielle Differentiation von (8.1) bezüglich der endogenen Variablen. Den absoluten Wert der Determinante dieser Differentiationen bezeichnen[3] wir mit $|\det J|$. Dies ist die Determinante der Jacobischen Matrix.

---

1) Koopmans, Rubin u. Leipnik, S. 71-72.

2) Vgl. auch Lüdeke, D.: Schätzprobleme in der Ökonometrie. Würzburg 1964, S. 22.

3) Die beiden senkrechten Striche geben an, daß es sich hier um den absoluten Wert der Determinante von J handelt.

Es gilt also[1)]

$$(8.6) \qquad \left|\frac{\partial u'}{\partial y}\right| = \det J,$$

wobei die linke Seite von (8.6) steht für

$$(8.7) \qquad \left|\frac{\partial u'}{\partial y}\right| = \begin{vmatrix} \frac{\partial u_1'}{\partial y_1} & \cdots & \frac{\partial u_T'}{\partial y_1} \\ \vdots & & \vdots \\ \frac{\partial u_1'}{\partial y_T} & \cdots & \frac{\partial u_T'}{\partial y_T} \end{vmatrix}$$

Die Ausdrücke unterhalb der Hauptdiagonalen dieser Determinanten sind gleich Null, da zukünftige Werte der endogenen Variablen die vorherigen latenten Variablen nicht beeinflussen können. Nun gilt aber bei Verwendung von (8.5) und (8.1) für die gleichzeitigen Kovarianzen

$$(8.8) \qquad \frac{\partial u_t'}{\partial y_t} = \frac{\partial(B\, y_t' + \Gamma\, z_t')}{\partial y_t} = B \qquad (t=1,\ldots,T)$$

Also erhalten wir unter Verwendung von (8.6), (8.7) und (8.8)

---

1) Vgl. Lüdeke: a.a.O., S. 22-23. Ohne die vorherige Notation umzuwerfen, sei vergegenwärtigt, daß

$$u' = \begin{bmatrix} u_{11} & \cdots & u_{1T} \\ \vdots & & \vdots \\ u_{G1} & \cdots & u_{GT} \end{bmatrix} \quad \text{und} \quad y = \begin{bmatrix} y_{11} & \cdots & y_{G1} \\ \vdots & & \vdots \\ y_{1T} & \cdots & y_{GT} \end{bmatrix},$$

durch Einsetzen von (t=1,...,T) in (8.2).

$$(8.9) \qquad \det J = \left|\frac{\partial u'}{\partial y}\right| = \begin{vmatrix} B & & & \\ 0 & B & & \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & B \end{vmatrix} = (\det B)^T$$

weil die Determinante von der Ordnung T ist und der Wert der Determinante bei triangulären Determinanten gleich dem Produkt der Elemente der Hauptdiagonalen ist.

Wir erhalten nun die Dichtefunktion der endogenen Variablen aus (8.5) durch

$$(8.10) \qquad p(y|z) = p(u) \cdot \left|\det \left(\frac{\partial u'}{\partial y}\right)\right| = p(u) \cdot \left|\det B\right|^T$$

Die Dichtefunktion der latenten Variablen wird also mit Hilfe der Jacobischen Determinante in die Dichtefunktion der endogenen Variablen transformiert.

Wir nehmen jetzt an[1)], daß die u_t (t=1,...,T) stochastisch unabhängig voneinander sind. Außerdem sollen die u_t für (t=1,...,T) einer gemeinsamen Dichtefunktion gehorchen. Letztere Annahme impliziert die zeitliche Konstanz des Verteilungsgesetzes u_t für (t=1,...,T).

Aus diesen Annahmen resultiert

$$(8.11) \qquad p(u) = \prod_{t=1}^{T} p(u_t)$$

(8.10) hat dann folgendes Aussehen

---

1) Vgl. Kapitel 1, Voraussetzungen 4 und 5.

$$(8.12) \qquad p(y|z) = \left|\det B\right|^T \prod_{t=1}^{T} p(u_t)$$

Schließlich wird angenommen[1], daß die u_t normalverteilt sind mit der Streuungs-Kovarianz-Matrix Σ. Es gilt also

$$(8.13) \qquad p(u_t) = \frac{1}{\sqrt{(2\pi)^G (\det \Sigma)}} \exp\left(-\frac{1}{2} u_t \Sigma^{-1} u_t'\right)$$

wobei der Skalar G für die Zahl der endogenen Variablen (bzw. Strukturgleichungen) steht. exp deutet an, daß es sich um eine Exponentialfunktion handelt. Daher ist beispielsweise $\exp a = e^a$.

Wir gelangen durch Einsetzen von (8.13) in (8.12) zu

$$(8.14) \qquad p(y|z) = \frac{\left|\det B\right|^T}{\sqrt{(2\pi)^{GT} (\det \Sigma)^T}} \exp\left[-\frac{1}{2} \sum_{t=1}^{T} (u_t \Sigma^{-1} u_t')\right]$$

(8.14) ist die Likelihoodfunktion, deren Maximum wir suchen. Wir logarithmieren nun die Likelihoodfunktion[2], dividieren dann mit T und erhalten die logarithmische Likelihoodfunktion

$$(8.15) \qquad L = \frac{1}{T} \ln p(y|z) = -\frac{G}{2} \ln (2\pi) - \frac{1}{2} \ln (\det \Sigma) + \ln \left|\det B\right| - \frac{1}{2T} \sum_{t=1}^{T} (u_t \Sigma^{-1} u_t').$$

---

1) Vgl. Kapitel 1, Voraussetzung 8.

2) Die Logarithmierung der Likelihoodfunktion mit dem logarithmus naturalis erleichtert die Maximierung insbesondere der Exponentialfunktion ($\ln e^a = a$), ohne das Maximum selbst zu beeinflussen.

Für den letzten Ausdruck von (8.15) können wir wegen (8.3) auch schreiben

$$(8.16) \quad -\frac{1}{2\bar{T}} \sum_{t=1}^{T} (u_t \, \Sigma^{-1} \, u_t') = -\frac{1}{2\bar{T}} \sum_{t=1}^{T} (x_t \, A'\Sigma^{-1} \, A \, x_t').$$

Da die Summierung über t den Teil $A'\Sigma^{-1}A$ nicht erfaßt, stellt der Ausdruck $\sum_{t=1}^{T} x_t \, x_t'$ die Diagonalelemente der Matrix x x' dar, wenn $x = [x_1, \ldots, x_T]$. Die Summe der Diagonalelemente einer quadratischen Matrix nennt man die Spur (Sp) dieser Matrix[1].

Der letzte Ausdruck von (8.15) lautet dann $-\frac{1}{2\bar{T}} \mathrm{Sp}(xA'\Sigma^{-1}Ax')$. Da $\mathrm{Sp}(x \, A'\Sigma^{-1}A \, x') = \mathrm{Sp}(\Sigma^{-1}A \, x'x \, A')$ und weil $\frac{1}{T}(x'x) = M_{xx}$, das heißt, die Momentenmatrix der x_t ist, erhalten wir

$$(8.17) \quad L = -\frac{G}{2} \ln (2\pi) - \frac{1}{2} \ln (\det \Sigma) + \ln |\det B| - \frac{1}{2} \mathrm{Sp}(\Sigma^{-1}A \, M_{xx} \, A').$$

Mit (8.17) läßt sich zeigen, daß äquivalente Strukturen, das heißt, Strukturen, die sich durch lineare Transformationen aus unserer gegebenen Struktur (8.3) herleiten lassen, dieselbe logarithmische Likelihoodfunktion haben wie die gegebene Struktur. Wenn die äquivalente Struktur von (8.3) also lautet[2]

$$(8.18) \quad A^* \, x_t' = \Lambda \, A \, x_t' = \Lambda \, u_t' \qquad (t=1,\ldots,T)$$

mit

---

1) Vgl. Menges (1961): a.a.O., S. 224.

2) Λ soll eine reguläre Matrix der Ordnung G sein.

$$(8.19) \qquad \Sigma^* = E(\Lambda\, u_t'\, u_t\, \Lambda') = \Lambda\, \Sigma\, \Lambda' \qquad (t=1,\ldots,T),$$

so hat die logarithmische Likelihoodfunktion von (8.18) folgendes Aussehen

$$(8.20) \quad L^* = -\frac{G}{2} \ln(2\pi) - 2 \cdot \frac{1}{2} \ln |\det \Lambda| - \frac{1}{2} \ln (\det \Sigma)$$
$$+ \ln |\det \Lambda| + \ln |\det B| - \frac{1}{2} \mathrm{Sp}[\Lambda^{-1}\Sigma^{-1}(\Lambda')^{-1} \cdot \Lambda A M_{xx} A' \Lambda']$$
$$= L,$$

da die Ausdrücke mit Λ sich gegeneinander aufheben[1]. Das Auffinden der "wahren" Struktur mit Hilfe der logarithmischen Likelihoodfunktion wird daher nur gewährleistet, wenn die "wahre" Struktur identifizierbar, also exakt oder überidentifiziert ist. Übrigens kann man auch zeigen, daß die logarithmische Likelihoodfunktion der reduzierten Form mit der der Struktur übereinstimmt, wenn man für Λ die Matrix B^{-1} setzt.

Um das Maximum von (8.17) zu finden, differenzieren wir (8.17) partiell nach Σ und setzen diese erste Ableitungen gleich Null. Wir erhalten[2]

$$(8.21) \quad \frac{\partial L}{\partial \hat{\Sigma}} = -\frac{1}{2} \mathrm{Sp}\left(\hat{\Sigma}^{-1} \frac{\partial \hat{\Sigma}}{\partial \hat{\Sigma}}\right) + \frac{1}{2} \mathrm{Sp}\left(\hat{\Sigma}^{-1} \frac{\partial \hat{\Sigma}}{\partial \hat{\Sigma}} \hat{\Sigma}^{-1} A\, M_{xx}\, A'\right)$$
$$= -\frac{1}{2} \mathrm{Sp}(\hat{\Sigma}^{-1}) + \frac{1}{2} \mathrm{Sp}(\hat{\Sigma}^{-1}\, \hat{\Sigma}^{-1}\, A\, M_{xx}\, A') = 0.$$

---

1) det Λ steht in dem zweiten Ausdruck auf der rechten Seite von (8.20) in Absolutzeichen, weil Λ in dem Ausdruck $-\frac{1}{2} \ln(\det \Lambda\, \Sigma\, \Lambda')$ quadratisch vorkommt.

2) Vgl. die Regeln für die Differentiation bei der Matrizenrechnung in: Menges (1961): a.a.O., S. 225.
Es ist zu beachten, daß Σ eine symmetrische Matrix ist, d.h. $\Sigma = \Sigma'$,

Die Bedingung, daß (8.21) gleich Null ist, wird dann erfüllt sein, wenn[1)]

$$(8.22) \qquad \hat{\Sigma} = A\, M_{xx}\, A' = \frac{1}{T}\,(u_t\, u_t')$$

Wenn wir (8.22) in (8.17) einsetzen, erhalten wir

$$(8.23) \qquad L(A) = -\frac{G}{2}(1 + \ln 2\pi) - \frac{1}{2}\ln(\det A\, M_{xx} A') + \ln|\det B|,$$

weil

$$(8.24) \qquad Sp[A\, M_{xx}\, A')^{-1}\, A\, M_{xx}\, A'] = Sp(I_G) = G.$$

(8.23) ist die durch Einsetzen der Schätzung der Streuungs-Kovarianz-Matrix konzentrierte Likelihoodfunktion, die nur noch bezüglich der Parametermatrix A[2)] maximiert wird. Hierbei erhalten wir[3)]

$$(8.25) \qquad \frac{\partial L(A)}{\partial A} = -\frac{1}{2}\, Sp[2(A\, M_{xx} A')^{-1}\, \frac{\partial A\, M_{xx} A'}{\partial A}] + Sp(B^{-1}\, \frac{\partial B}{\partial A})$$

$$= -\, Sp[(A\, M_{xx}\, A')^{-1}\, A\, M_{xx}] + Sp(B^{-1}\, \frac{\partial B}{\partial A}) = 0,$$

da im Maximum die Steigung (1. Ableitung) gleich Null ist.[4)]

---

1) Vgl. auch Koopmans, Rubin u. Leipnik: a.a.O., S. 118.

2) Liegen a-priori-Restriktionen über A vor, so wird nur bezüglich des Teils von A maximiert, der freie Parameter enthält. Insoweit ist bei der Maximierung nur der Teil von (8.23) relevant, der die Untermatrix der freien Parameter enthält.

3) Vgl. Menges (1961): a.a.O., S. 225 u. S. 224.

4) Chernoff und Divinsky unterscheiden den Diagonalfall und den Nicht-Diagonalfall der Streuungs-Kovarianz-Matrix der latenten Variablen und entwickeln für beide Fälle Iterationsverfahren. Vgl. Chernoff, H. u. Divinsky, N.: The computation of maximum-likelihood estimates of linear structural equations. In: Studies in Econometric Method. Hrsg.: W.C. Hood u. Tjalling C. Koopmans. New York-London 1953, S. 236-269, s. bes. S. 252-259 u. S. 263 - 267.

Wenn wir die Matrix A in ihre Untermatrizen B und Γ zerlegen, stellen wir fest, daß der zweite Ausdruck auf der rechten Seite von (8.25) verschwindet, wenn nach den Elementen von Γ differenziert wird, da diese nicht in B enthalten sind.

Für die reduzierte Form ergeben sich weitere Vereinfachungen, da dann für B I_G gesetzt werden kann und A folgendes Aussehen hat: $A = [I_G \quad \Gamma]$.

Auch bei einer einzigen Gleichung ergeben sich Vereinfachungen, weil $(A\, M_{xx}\, A')^{-1}$ dann ein Skalar ist.

Wir wollen noch die Größte-Dichte-Schätzung für den Diagonalfall[1] untersuchen. Der Diagonalfall liegt dann vor, wenn Autokorrelation nicht zugelassen ist. Die Streuungs-Kovarianz-Matrix der latenten Variablen hat folgendes Aussehen

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \sigma_{GG}^2 \end{bmatrix} . \tag{8.26}$$

Wieder setzen wir in die logarithmische Likelihoodfunktion (8.15) ein und erhalten

$$L_D = -\frac{G}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{G} \ln \sigma_{ii}^2 + \ln|\det B| \tag{8.27}$$

$$= \frac{1}{2T}\sum_{t=1}^{T}\sum_{i=1}^{G}\frac{u_{it}^2}{\sigma_{ii}^2}$$

1) Vgl. Menges (1961): a.a.O., S. 101-102 u. die Differentiationsregeln auf S. 225.
Vgl. auch Kapitel 1, Voraussetzungen 7 und 8.

Um das Maximum von (8.27) zu finden, differenzieren wir (8.27) partiell elementweise nach den Parametern σ_{ii} und a_{ij} und setzen diese ersten Ableitungen gleich Null. Wir erhalten

$$(8.28) \qquad \frac{\partial L_D}{\partial \hat{\sigma}_{ii}} = -\frac{1}{\hat{\sigma}_{ii}} + \frac{1}{T\,\hat{\theta}_{ii}^3} \sum_{t=1}^{T} u_{it}^2 = 0$$

und

$$(8.29) \qquad \frac{\partial L_D}{\partial \hat{a}_{ij}} = Sp(B^{-1} \frac{\partial B}{\partial \hat{a}_{ij}}) - \frac{1}{T\,\hat{\sigma}_{ii}^2} \sum_{t=1}^{T} \frac{\partial u_{it}^2}{\partial \hat{a}_{ij}} = 0.$$

Eine Umformung von (8.28) ergibt

$$(8.30) \qquad \hat{\sigma}_{ii}^2 = \frac{1}{T} \sum_{t=1}^{T} u_{it}^2 .$$

Eine Umformung von (8.29) ergibt

$$(8.31) \qquad \frac{1}{T} \sum_{t=1}^{T} u_{it}\, x_j = \hat{\sigma}_{ii}^2\, Sp(B^{-1} \frac{\partial B}{\partial \hat{a}_{ij}})$$

(8.31) geht über in die Normalgleichungen, die mit der Methode der kleinsten Quadrate gewonnen werden, wenn in (8.27) det B = 1. Dies ist entweder der Fall, wenn unser Modell (8.1) in der reduzierten Form geschrieben werden kann, weil dann B gleich der Einheitsmatrix ist. det B = 1 gilt jedoch auch dann, wenn die Matrix B triangulär ist und die Parameter β_{ii} auf der Hauptdiagonalen gleich 1 sind, also die β_{ii} normiert sind. In diesem zweiten Fall hat B folgende Gestalt[1)]

---

1) Auch die Anordnung B' einer triangulären Matrix führt zum gleichen Ergebnis.

$$(8.32) \qquad B = \begin{bmatrix} 1 & \beta_{12} & \cdots & \beta_{1G} \\ 0 & 1 & \cdots & \beta_{2G} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Eine solche Parametermatrix kommt bei rekursiven Systemen vor.

Die Schätzungen nach der Methode der größten Dichte bei vollständiger Information sind konsistente[1] Schätzungen. Unter ziemlich verallgemeinerten Bedingungen[2] sind diese Schätzungen asymptotisch normal und asymptotisch effizient.

---

1) Vgl. Cramér: a.a.O., S. 500.
Für Spezialfälle wie etwa Quasi-Größte-Dichte-Schätzungen (Diese liegen vor, wenn nicht von der Normalverteilungsvoraussetzung ausgegangen wird.) vgl.:
Mann, H.B. u. Wald, A.: On the statistical treatment of linear stochastic difference equations. Econometrica, 11(1943), S. 173-220, s. bes. S. 203-204.
Koopmans, Rubin u. Leipnik: a.a.O., S. 133-153.
Rubin, H.: Consistency of maximum-likelihood estimates in the explosive case. In: Statistical Inference in Dynamic Economic Models. Hrsg.: Tjalling C. Koopmans. New York-London 1950, S. 356-364.

2) Vgl. hierzu Cramér: a.a.O., Abschnitt 33.3, s. bes. S. 500.

9. Die Methode der größten Dichte bei beschränkter Information nach Anderson und Rubin[1)]

Die Methode der größten Dichte bei beschränkter Information (limited-information maximum-likelihood method) wurde erstmals 1949 von Theodore W. Anderson, Jr. und Herman Rubin publiziert in den Annals of Mathematical Statistics. Diese Methode ist zur Schätzung der Parameter einer einzigen Gleichung aus einem vollständigen Gleichungssystem geeignet, ohne daß die Parameter der restlichen Gleichungen des Systems geschätzt werden müssen. Erforderlich für diese Methode ist nur eine Aufstellung der aus dieser Gleichung ausgeschlossenen exogenen Variablen des Systems, m.a.W. müssen a-priori-Restriktionen (Nullrestriktionen) über die Parameter dieser ausgeschlossenen Variablen vorliegen. Außerdem muß die reduzierte Form berechnet werden können, damit die reduzierte-Form-Parameter geschätzt werden können.

Wir gehen von folgendem Gleichungssystem von Strukturgleichungen aus[2)3)]

---

1) Anderson, T.W. u. Rubin, H.: Estimation of the parameters of a single equation in a complete system of stochastic equations. The Annals of Mathematical Statistics, 20(1949), S. 46-63.

Anderson, T.W. u. Rubin, H.: The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations. The Annals of Mathematical Statistics, 21(1950), S. 570-582.

Vgl. auch Anderson, T.W.: Estimation of the parameters of a single equation by the limited-information maximum-likelihood method. In: Statistical Inference in Dynamic Economic Models. Hrsg.: Tjalling C. Koopmans. New York-London 1950, S. 311-322.

2) Vgl. Johnston: a.a.O., S. 254-258 u. S. 240.

3) Wieder werden die Skalare besonders hervorgehoben.

$$(9.1) \qquad B\, y_t + \Gamma\, x_t = u_t \qquad (t=1,\ldots,T),$$

wobei

B = die Matrix der Koeffizienten der gemeinsam abhängigen Variablen der Ordnung G,

Γ = die Matrix der Koeffizienten der vorherbestimmten Variablen der Ordnung G x K;

die Vektoren y_t, x_t und u_t enthalten folgende Skalare

$$(9.2) \qquad y_t = \begin{bmatrix} y_{1t} \\ \vdots \\ y_{Gt} \end{bmatrix}, \quad x_t = \begin{bmatrix} x_{1t} \\ \vdots \\ x_{Kt} \end{bmatrix} \quad \text{und} \quad u_t = \begin{bmatrix} u_{1t} \\ \vdots \\ u_{Gt} \end{bmatrix}.$$

Wir betrachten die erste Gleichung des Systems (9.1)

$$(9.3) \qquad \beta_1\, y_t + \gamma_1\, x_t = u_{1t} \qquad (t=1,\ldots,T)$$

wobei β_1 die erste Zeile von B und γ_1 die erste Zeile von Γ sind. Durch die Verwendung der a-priori-(Null-) Restriktionen der Koeffizienten von (9.3) erhalten wir in Skalarschreibweise

$$(9.4) \qquad \beta_{11}\, y_{1t} + \ldots + \beta_{1G}\Delta y_G \Delta_t + \gamma_{11}\, x_{1t} + \ldots + \gamma_{1K^*}\, x_{K^*t} = u_{1t} \qquad (t=1,\ldots,T).$$

Die reduzierte Form des Systems (9.1) lautet nun

$$(9.5) \qquad y_t = \Pi\, x_t + v_t \qquad (t=1,\ldots,T)$$

wobei Π = die Matrix der reduzierte-Form-Koeffizienten der Ordnung G x K

und v_t = der Spaltenvektor der G reduzierte-Form-Störvariablen sind.

Diese reduzierte Form erhalten wir dann und nur dann, wenn wir (9.1) von links mit der Kehrmatrix von B, nämlich B^{-1}, multiplizieren können. B^{-1} existiert jedoch dann und nur dann, wenn B regulär ist, d.h. wenn det $B \neq 0$. Diese Bedingung ist verletzt, wenn in (9.1) Definitionsgleichungen enthalten sind. Daher müssen Definitionsgleichungen in die eigentlichen Strukturgleichungen (welche latente Variablen aufweisen) eingesetzt werden, um die Regularität von B zu gewährleisten. Wenn wir (9.1) von links mit B^{-1} multiplizieren, erhalten wir

$$(9.6) \qquad y_t = - B^{-1} \Gamma x_t + B^{-1} u_t \qquad (t=1,\ldots,T).$$

Durch die Identität von (9.6) und (9.5) ergeben sich

$$(9.7) \qquad \Pi = - B^{-1} \Gamma$$

und

$$(9.8) \qquad v_t = B^{-1} u_t \qquad (t=1,\ldots,T).$$

Wir formen (9.7) um durch Multiplikation mit B von links

$$(9.9) \qquad - \Gamma = B \Pi.$$

Die erste Gleichung dieses Systems (9.9) lautet dann

$$(9.10) \qquad - \gamma_1 = \beta_1 \Pi$$

wobei γ_1 und β_1 jeweils in zwei Untervektoren zerlegt[1] werden können. Berücksichtigt man, daß γ_1 K Skalare und β_1 G Skalare enthalten, so zeigt sich

---

1) Wenn wir die a-priori-(Null-) Restriktionen berücksichtigen.

(9.11) $\gamma_1 = [\gamma_{1*} \quad 0_{1**}] = [\gamma_{11} \cdots \gamma_{1K^*} \; 0_{1,K^*+1} \cdots 0_{1K}]$,

(9.12) $\beta_1 = [\beta_{1\Delta} \quad 0_{1\Delta\Delta}] = [\beta_{11} \cdots \beta_{1G^\Delta} \; 0_{1,G^\Delta+1} \cdots 0_{1G}]$

und entsprechend

$$(9.13) \qquad \Pi = \begin{bmatrix} \Pi_{\Delta^*} & \Pi_{\Delta,**} \\ \Pi_{\Delta\Delta,*} & \Pi_{\Delta\Delta,**} \end{bmatrix}.$$

(9.10) läßt sich dann wie folgt schreiben:

$$(9.14) \qquad -[\gamma_{1*} \quad 0_{1**}] = [\beta_{1\Delta} \quad 0_{1\Delta\Delta}] \begin{bmatrix} \Pi_{\Delta^*} & \Pi_{\Delta,**} \\ \Pi_{\Delta\Delta,*} & \Pi_{\Delta\Delta,**} \end{bmatrix}.$$

Hieraus folgt

$$(9.15) \qquad -\gamma_{1*} = \beta_{1\Delta} \Pi_{\Delta^*}$$

und

$$(9.16) \qquad 0_{1**} = \beta_{1\Delta} \Pi_{\Delta,**} .$$

Für (9.16) können wir schreiben:

$$(9.17) \qquad [0_{1,K^*+1} \cdots 0_{1K}] = [\beta_{11} \cdots \beta_{1G^\Delta}] \begin{bmatrix} \Pi_{1,K^*+1} & \cdots & \Pi_{1,K^{**}} \\ \vdots & & \vdots \\ \Pi_{G^\Delta,K^*+1} & \cdots & \Pi_{G^\Delta,K^{**}} \end{bmatrix}.$$

Die elementarste Bedingung für die Auflösung von (9.17) nach den Strukturparametern $[\beta_{11} \cdots \beta_{1G^\Delta}]$ besteht darin, daß die reduzierte-Form-Parameter Π, insbesondere $\Pi_{\Delta,**}$, überhaupt geschätzt werden können.

Die uns besonders interessierenden ersten G^Δ Gleichungen des reduzierte-Form-Gleichungssystems können eventuell

geschätzt werden, wenn das Abzählkriterium[1] für Identifikation erfüllt ist, d.h. wenn

$$K^{**} \geq G^{\Delta} - 1. \tag{9.18}$$

Zuverlässiger ist das Rangkriterium[2] für Identifikation, weil es die Möglichkeit von Linearkombinationen der Gleichungen berücksichtigt. Nach dem Rangkriterium können die ersten G^{Δ} Gleichungen des reduzierte-Form-Gleichungssystems dann und nur dann geschätzt werden, wenn der Rang von $\Pi_{\Delta,**}$ mindestens gleich $(G^{\Delta} - 1)$ ist.

Ist der Rang von $\Pi_{\Delta,**}$ kleiner als $(G^{\Delta} - 1)$, so liegt Unteridentifikation vor. Dann lassen sich keine reduzierte-Form-Parameterschätzungen $\hat{\Pi}_{\Delta,**}$ finden[3]. Kennt man $\hat{\Pi}_{\Delta,**}$ nicht, so läßt sich auch das Gleichungssystem

$$O_{1**} = \hat{B}_{1\Delta}\,\hat{\Pi}_{\Delta,**} \tag{9.19}$$

nicht nach dem gesuchten $\hat{B}_{1\Delta}$ auflösen.

Wir wollen uns nun dem Fall der exakten Identifikation zuwenden, der dann vorliegt, wenn die Bedingung (9.18) durch

$$K^{**} = G^{\Delta} - 1 \tag{9.20}$$

erfüllt ist[4]. Die reduzierte-Form-Parameter lassen sich also schätzen, so daß uns für das Gleichungssystem (9.17) der Zeilenvektor der Null-Koeffizienten und die Matrix

---

1) Vgl. Menges (1961): a.a.O., S. 82-83.

2) Vgl. Menges (1961): a.a.O., S. 82.

3) Eigentlich läßt sich bei Unteridentifikation die reduzierte Form gar nicht erst berechnen, weil B infolge der Linearkombinationen der Strukturgleichungen nicht mehr regulär ist.

4) Nach dem Rangkriterium muß der Rang von $\Pi_{\Delta,**}$ genau gleich $(G^{\Delta} - 1)$ sein.

der reduzierte-Form-Parameter bekannt sind, der Zeilenvektor $[\beta_{11} \ldots \beta_{1G}]$ jedoch unbekannt ist.

Da wir einen unbekannten Parameter β_{1i} mehr haben, als uns Gleichungen zur Verfügung stehen, handelt es sich um ein unterbestimmtes Gleichungssystem, das aber, weil nur ein Freiheitsgrad ($G^{\Delta} - K^{**} = 1$) vorliegt, durch Normierung gelöst werden kann. M.a.W., wenn wir einen der Koeffizienten β_{1i}, beispielsweise β_{11}, Eins gleichsetzen, so erhalten wir

$$(9.21) \qquad \frac{\beta_{11}}{\beta_{11}} = 1, \quad \frac{\beta_{12}}{\beta_{11}} = \beta_{12}, \ \ldots, \ \frac{\beta_{1G^{\Delta}}}{\beta_{11}} = \beta_{1G^{\Delta}} \quad (\text{für } \beta_{11} \neq 0)$$

Wir erhalten also für alle Elemente dieses $\beta_{1\Delta}$-Vektors ihren genauen Wert, wenn wir bedenken, daß sich bei der Auflösung der ersten Gleichung (9.3) von (9.1) nach y_{1t} ergibt $\beta_{11} = 1$. Diese für den Fall der exakten Identifikation beschriebene Verfahrensweise (Berechnung der reduzierten Form, Schätzung der reduzierte-Form-Parameter, Auflösen nach den Parameterschätzungen der Strukturgleichung) entspricht genau dem Vorgehen der indirekten Methode der kleinsten Quadrate.

Wir betrachten jetzt den Fall der Überidentifikation. Dieser tritt ein, wenn (9.18) durch

$$(9.22) \qquad K^{**} > G^{\Delta} - 1$$

erfüllt ist, d.h. wenn $K^{**} \geq G^{\Delta}$. Überidentifikation liegt also vor, wenn es mindestens so viele Gleichungen (nämlich K^{**}) wie unbekannte Parameter β_{1i} (nämlich G^{Δ}) im Gleichungssystem (9.17) gibt.[1] Da (9.17) ein homo-

1) Hier zeigt uns das Rangkriterium, daß bei Überidentifikation nur der Fall $\rho(\Pi_{\Delta,**}) = G^{\Delta}$ relevant ist, da der Rang von $\Pi_{\Delta,**}$ höchstens gleich G^{Δ} ist, weil $\Pi_{\Delta,**}$ nur aus G^{Δ} Zeilen besteht. Deshalb kann ein höherer Rang als G^{Δ} nicht vorkommen.

genes Gleichungssystem[1] ist, kann in diesem Fall der Lösungsvektor nur der Nullvektor sein, es ist also nur eine triviale Lösung unseres Gleichungssystems möglich. Dies tritt also auch dann ein, wenn das Gleichungssystem genau bestimmt ist, es also genau so viele Gleichungen wie endogene Variablen gibt ($K^{**} = G^{\Delta}$). Der Veranschaulichung halber wollen wir diesen interessanten Unterfall an einem homogenen Gleichungssystem mit zwei Gleichungen und zwei endogenen Variablen darstellen:

$$(9.23) \qquad \begin{aligned} 0 &= \Pi_{11}\, \beta_1 + \Pi_{12}\, \beta_2 \\ 0 &= \Pi_{21}\, \beta_1 + \Pi_{22}\, \beta_2, \end{aligned}$$

wobei β_1 und β_2 die Unbekannten sind und Π_{11}, Π_{12}, Π_{21} und Π_{22} als bekannt vorausgesetzt werden.

Die Lösung nach der Cramérschen Regel[2] lautet dann

$$(9.24) \qquad \beta_1 = \frac{\begin{vmatrix} 0 & \Pi_{12} \\ 0 & \Pi_{22} \end{vmatrix}}{\begin{vmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{vmatrix}} \quad \text{und} \quad \beta_2 = \frac{\begin{vmatrix} \Pi_{11} & 0 \\ \Pi_{21} & 0 \end{vmatrix}}{\begin{vmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{vmatrix}}.$$

Da der Wert der Determinanten im Zähler gleich Null ist, weil jeweils eine Spalte nur aus Nullen besteht, gilt für dieses homogene Gleichungssystem

$$(9.25) \qquad \beta_1 = \beta_2 = 0,$$

wenn die Hauptdeterminante von Null verschieden ist.

---

1) Vgl. auch Menges (1961): a.a.O., S. 223-224. Dort wird das Problem in aller mathematischer Genauigkeit abgehandelt.

2) Vgl. Zurmühl: a.a.O., S. 39.

Dies läßt sich auch für homogene Gleichungssysteme mit n Gleichungen und n endogenen Variablen zeigen, da nach der Cramérschen Regel in jedem Fall im Zähler eine Determinante steht, bei der in einer Spalte nur Nullen enthalten sind.[1)]

Zusammenfassend können wir feststellen, daß es eine Lösung von (9.16) nur dann gibt, wenn exakte Identifikation vorliegt[2)], wenn also (9.20) gilt.

Durch Einsetzen des mit (9.16) gefundenen Lösungsvektors $\beta_{1\Delta}$ in (9.15) läßt sich sodann der Lösungsvektor γ_{1*} ermitteln.

Die Methode der größten Dichte bei beschränkter Information ist im Gegensatz zur indirekten Methode der kleinsten Quadrate auch anwendbar, wenn Überidentifikation vorliegt. Bei der folgenden Darstellung der von Anderson und Rubin entwickelten Methode der größten Dichte bei beschränkter Information gehen wir davon aus, daß (9.3) identifizierbar ist (Überidentifikation bzw. exakte Identifikation). Außerdem sei B regulär.

Wir verwenden statt der etwas verwirrenden Anderson-Rubin-Symbole[3)] die bisherigen Symbole dieses Kapitels (Johnston-Symbole) und zerlegen nun (9.3) aufgrund der a-priori-Restriktionen bereits vor der Gewinnung der reduzierten Form.

---

1) Vgl. auch Menges (1961): a.a.O., S. 223.

2) Die Methode der größten Dichte bei beschränkter Information reduziert bei Überidentifikation den Rang von $\Pi_{\Delta,**}$ von G^{Δ} auf $G^{\Delta} - 1$ durch eine zusätzliche Restriktion: Die durch x_{t**} erklärte Streuung soll ein minimales Bruchteil der Streuung der latenten Variablen der reduzierten Form sein.

3) Anderson u. Rubin (1949): a.a.O., S. 48-49.

Wir gehen also aus von der Strukturgleichung[1)]

$$(9.3) \qquad \beta_1 y_t + \gamma_1 x_t = u_{1t} \qquad (t=1,\ldots,T)^{2)}$$

und zerlegen die Vektoren von (9.3) in

$$(9.26) \qquad \gamma_1 = [\gamma_{1*} \quad 0_{1**}]$$

$$(9.27) \qquad \beta_1 = [\beta_{1\Delta} \quad 0_{1\Delta\Delta}]$$

$$(9.28) \qquad x_t = \begin{bmatrix} x_{t*} \\ x_{t**} \end{bmatrix}$$

und

$$(9.29) \qquad y_t = \begin{bmatrix} y_{t_\Delta} \\ y_{t_{\Delta\Delta}} \end{bmatrix}.$$

Dann erhalten wir aus (9.3) durch Einsetzen der Zerlegungen und Ausmultiplikation

$$(9.30) \qquad \beta_{1\Delta} y_{t\Delta} + 0_{1\Delta\Delta} y_{t\Delta\Delta} + \gamma_{1*} x_{t*} + 0_{1**} x_{t**} = u_{1t},$$

was wir wegen der Null-Restriktionen auch wie folgt schreiben können

$$(9.31) \qquad \beta_{1\Delta} y_{t\Delta} + \gamma_{1*} x_{t*} = u_{1t}.$$

Wenn wir in die reduzierte Form

$$(9.5) \qquad y_t = \Pi x_t + v_t$$

---

1) Vgl. hierzu Anderson u. Rubin (1949): a.a.O., S. 48-50.

2) Der Einfachheit halber soll bei der folgenden Darstellung (t=1,...,T) nicht mitgeführt werden; gleichwohl soll der mitgeführte Index t die Möglichkeit der Zusammenfassung der Beobachtungswerte zu Vektoren und Matrizen vergegenwärtigen.

die Zerlegungen (9.13), (9.28), (9.29) und

$$(9.32) \qquad v_t = \begin{bmatrix} \delta_t \\ \zeta_t \end{bmatrix}$$

einsetzen und dann ausmultiplizieren, so gelangen wir zu

$$(9.33) \qquad \begin{bmatrix} y_{t\Delta} \\ y_{t\Delta\Delta} \end{bmatrix} = \begin{bmatrix} \Pi_{\Delta^*} x_{t*} + \Pi_{\Delta,**} x_{t**} + \delta_t \\ \Pi_{\Delta\Delta,*} x_{t*} + \Pi_{\Delta\Delta,**} x_{t**} + \zeta_t \end{bmatrix}.$$

Durch Multiplikation von (9.33) mit $B_1 = [B_{1\Delta} \quad 0_{1\Delta\Delta}]$ von links erhalten wir bei Berücksichtigung der Nullrestriktionen

$$(9.34) \qquad B_{1\Delta} y_{t\Delta} = B_{1\Delta} \Pi_{\Delta^*} x_{t*} + B_{1\Delta} \Pi_{\Delta,**} x_{t**} + B_{1\Delta} \delta_t .$$

Der Vergleich von (9.34) mit (9.31) ergibt schließlich die bereits bekannten Relationen

$$(9.15) \qquad -\gamma_{1*} = B_{1\Delta} \Pi_{\Delta^*}$$

und

$$(9.16) \qquad 0_{1**} = B_{1\Delta} \Pi_{\Delta,**}$$

sowie

$$(9.35) \qquad u_{1t} = B_{1\Delta} \delta_t .$$

Aus (9.35) erhalten wir weiter

$$(9.36) \qquad u_{1t} u_{1t}' = B_{1\Delta} \delta_t \delta_t' B_{1\Delta}' .$$

Auf der Grundlage von (9.36) läßt sich die Schätzung der Streuungs-Kovarianz-Matrix der latenten Variablen der

Strukturgleichung (9.3) mit Hilfe des geschätzten Parametervektors $\hat{\beta}_{1\Delta}$, und der Schätzung der latenten Variablen der ersten $G^{\wedge}$ Gleichungen der reduzierten Form, $\hat{\delta}_t$, berechnen.

Wir betrachten nun die obere Gleichung aus (9.33)

$$(9.37) \qquad y_{t\Delta} = \Pi_{\Delta *}\, x_{t*} + \Pi_{\Delta,**}\, x_{t**} + \delta_t$$

etwas genauer.

$y_{t\Delta}$ hängt also bei der reduzierte-Form-Schätzung im wesentlichen von zwei Komponenten ab, nämlich von den in der Strukturgleichung (9.31) enthaltenen vorherbestimmten Variablen x_{t*} einerseits und von den aus (9.31) ausgeschlossenen vorherbestimmten Variablen x_{t**} andererseits. Um die Möglichkeit der Kollinearität zwischen diesen beiden Gruppen vorherbestimmter Variablen auszuschalten, schätzen Anderson und Rubin x_{t**} aufgrund von x_{t*} und verwenden im nachfolgenden Schätzvorgang statt x_{t**} nur noch den von x_{t*} unabhängigen Rest $x_{t**} - \hat{x}_{t**} = \tilde{x}_{t**}$.

Die Kleinst-Quadrat-Schätzung von x_{t**} lautet[1)]

$$(9.38) \qquad \hat{x}_{t**} = x_{t**}\, x'_{t*}(x_{t*}\, x'_{t*})^{-1}\, x_{t*}.$$

Dann ist unsere "neue" Variable

$$(9.39) \qquad \tilde{x}_{t**} = x_{t**} - x_{t**}\, x'_{t*}(x_{t*}\, x'_{t*})^{-1}\, x_{t*}.$$

---

1) Vgl. (2.30). Die abweichende Schreibweise ergibt sich wegen der transponierten Gestalt der Vektoren x_{t*} im Vergleich zu den Zeilen von X (vgl. (2.17)) und wegen des Voranstellens der Parametervektoren.

Statt (9.37) verwenden wir im folgenden

$$y_{t\Delta} = \tilde{\Pi}_{\Delta *}\, x_{t*} + \tilde{\Pi}_{\Delta,**}\, \tilde{x}_{t**} + \delta_t, \tag{9.40}$$

wobei

$$\tilde{\Pi}_{\Delta *} = \Pi_{\Delta *} + \Pi_{\Delta,**}\, x_{t**}\, x'_{t*}(x_{t*}\, x'_{t*})^{-1} \tag{9.41}$$

und

$$\tilde{\Pi}_{\Delta,**} = \Pi_{\Delta,**}\ . \tag{9.42}$$

Durch Einsetzen von (9.41) und (9.42) in (9.40) kann man zeigen, daß (9.40) eigentlich nur eine algebraische Umformung von (9.37) ist.[1)]

Wir betrachten hier noch einen anderen wichtigen Aspekt. Wenn wir (9.42) in (9.40) einsetzen und dann beide Seiten der neuen Gleichung von links mit $\beta_{1\Delta}$ multiplizieren, so erhalten wir

$$\beta_{1\Delta} y_{t\Delta} = \beta_{1\Delta}\tilde{\Pi}_{\Delta *} x_{t*} + \beta_{1\Delta}\Pi_{\Delta,**}\tilde{x}_{t**} + \beta_{1\Delta}\delta_t\ . \tag{9.43}$$

Durch Einsetzen von (9.16) und (9.35) in (9.43) erhalten wir mit

$$\beta_{1\Delta} y_{t\Delta} = \beta_{1\Delta}\tilde{\Pi}_{\Delta *}\, x_{t*} + u_{1t} \tag{9.44}$$

---

1) Wir erhalten, ausgehend von (9.40),

$$\begin{aligned} y_{t\Delta} &= \Pi_{\Delta *}\, x_{t*} + \Pi_{\Delta,**}\, x_{t**}\, x'_{t*}(x_{t*}\, x'_{t*})^{-1}\, x_{t*} + \Pi_{\Delta,**} x_{t**} \\ &\quad - \Pi_{\Delta,**}\, x_{t**}\, x'_{t**}(x_{t*}\, x'_{t*})^{-1}\, x_{t*} + \delta_t \\ &= \Pi_{\Delta *}\, x_{t*} + \Pi_{\Delta,**}\, x_{t**} + \delta_t = (9.37). \end{aligned}$$

unsere Strukturgleichung, in der x_{t**} und damit auch $\tilde{x}_{t**}$ a priori ausgeschlossen sind. Daß (9.44) und (9.31) identisch sind, läßt sich durch Einsetzen von (9.41) in (9.44) und wiederum durch Verwendung von (9.15) und (9.16) beweisen. Dabei zeigt sich, daß

$$(9.45) \qquad \beta_{1\Delta}\,\tilde{\Pi}_{\Delta *} = -\gamma_{1*}.$$

Durch Einsetzen von (9.42) in (9.16) erhalten wir noch

$$(9.46) \qquad 0_{1**} = \beta_{1\Delta}\,\tilde{\Pi}_{\Delta,**}.$$

Wir sehen, daß auch für die "neuen" reduzierte-Form-Parameter analoge Bedingungen gelten wie für die früheren reduzierte-Form-Parameter (Vgl. mit (9.15) und (9.16)).

Die Schätzung von (9.40) mit der Methode der kleinsten Quadrate ergibt folgende Parameterschätzungen

$$(9.47) \qquad \hat{\Pi}_{\Delta *} = y_{t_\Delta}\, x'_{t*}\,(x_{t*}\, x'_{t*})^{-1}$$

und

$$(9.48) \qquad \hat{\Pi}_{\Delta,**} = y_{t_\Delta}\, \tilde{x}'_{t**}(\tilde{x}_{t**}\, \tilde{x}'_{t**})^{-1}.$$

Ferner berechnen wir mit Hilfe von (9.40)

$$(9.49) \quad \delta_t\, \delta'_t = [I - \tilde{\Pi}_{\Delta *} - \tilde{\Pi}_{\Delta,**}]\begin{bmatrix} y_{t_\Delta} \\ x_{t*} \\ \tilde{x}_{t**} \end{bmatrix}[y'_{t\Delta}\; x'_{t*}\; \tilde{x}'_{t**}]\begin{bmatrix} I \\ -\tilde{\Pi}'_{\Delta *} \\ -\tilde{\Pi}'_{\Delta,**} \end{bmatrix}.$$

Durch Ausmultiplikation erhält man[1)]

---

1) Es ist zu beachten, daß $\tilde{x}_{t**}$ ein Residuum ist, welches orthogonal zu x_{t*} ist. Bei den restlichen Ausdrücken ist auf Symmetrien zu achten.

$$(9.50)\quad \delta_t \delta_t' = y_{t\Delta} y_{t\Delta}' - \tilde{\Pi}_{\Delta *} x_{t*} x_{t*}' \tilde{\Pi}_{\Delta *}' - \tilde{\Pi}_{\Delta,**} \tilde{x}_{t**} \tilde{x}_{t**}' \tilde{\Pi}_{\Delta,**}'.$$

Wegen (9.36) erhalten wir aus (9.50)

$$(9.51)\quad u_{1t} u_{1t}' = \beta_{1\Delta} y_{t\Delta} y_{t\Delta}' \beta_{1\Delta}' - \beta_{1\Delta} \bar{\Pi}_{\Delta *} x_{t*} x_{t*}' \tilde{\Pi}_{\Delta *}' \beta_{1\Delta}' - \beta_{1\Delta} \tilde{\Pi}_{\Delta,**} \tilde{x}_{t**} \tilde{x}_{t**}' \tilde{\Pi}_{\Delta,**}' \beta_{1\Delta}'.$$

Hieraus folgt wegen (9.46) und (9.45)

$$(9.52)\quad u_{1t} u_{1t}' = \beta_{1\Delta} y_{t\Delta} y_{t\Delta}' \beta_{1\Delta}' - \gamma_{1*} x_{t*} x_{t*}' \gamma_{1*}'.$$

So sehen wir, daß die durch $\tilde{x}_{t**}$ erklärte Steuung von $y_{t\Delta}$ in der Strukturgleichung aufgrund der a-priori-Restriktionen verschwindet.

Entsprechend soll auch bei der Schätzung von (9.51) der folgende Ausdruck

$$(9.53)\quad \hat{\beta}_{1\Delta} \hat{\Pi}_{\Delta,**} \tilde{x}_{t**} \tilde{x}_{t**}' \hat{\Pi}_{\Delta,**}' \hat{\beta}_{1\Delta}'$$

bezüglich $\hat{\beta}_{1\Delta}$ minimiert werden[1], da die Verminderung der Schätzung von $u_{1t} u_{1t}'$ durch den Ausdruck (9.53) den a-priori-Informationen von Gleichung (9.44) widerspricht. Gewissermaßen soll also dieser Widerspruch bei der Schätzung minimiert werden.

Da auch die Schätzung von $u_{1t} u_{1t}'$, nämlich[2]

$$(9.54)\quad \hat{u}_{1t} \hat{u}_{1t}' = \hat{\beta}_{1\Delta} \hat{\delta}_t \hat{\delta}_t' \hat{\beta}_{1\Delta}'$$

bezüglich $\hat{\beta}_{1\Delta}$ minimiert werden[3] muß, stehen wir vor einer komplizierten Minimumaufgabe:

---

1) (9.53) verschwindet nicht, solange $\hat{\Pi}_{\Delta,**} \neq \tilde{\Pi}_{\Delta,**}$ [Vgl. (9.46)].

2) Vgl. mit (9.36) und dann mit (9.50).

3) Hier soll die Summe der Quadrate der Residuen minimiert werden.

Sowohl (9.53) als auch (9.54) sind bezüglich $\hat{\beta}_{1\Delta}$ zu minimieren[1], wobei jedoch darauf zu achten ist, daß (9.53) das kleinste[2] Vielfache ν von (9.54) ist.

Es gilt also für die erste Ableitung bei der Minimierung von (9.53)

$$(9.55)\quad \frac{\partial \hat{\beta}_{1\Delta} \hat{\tilde{\Pi}}_{\Delta,**} \tilde{x}_{t**} \tilde{x}'_{t**} \hat{\tilde{\Pi}}'_{\Delta,**} \hat{\beta}'_{1\Delta}}{\partial \hat{\beta}_{1\wedge}} = \frac{\partial \nu \hat{\beta}_{1\wedge} \hat{\delta}_t \hat{\delta}'_t \hat{\beta}'_{1\Delta}}{\partial \hat{\beta}_{1\wedge}} = 0_{G\Delta}$$

wobei $0_{G\Delta}$ der Nullzeilenvektor mit $G^{\wedge}$ Elementen ist. Hieraus ergibt sich

$$(9.56)\quad \hat{\beta}_{1\Delta} \hat{\tilde{\Pi}}_{\Delta,**} \tilde{x}_{t**} \tilde{x}'_{t**} \hat{\tilde{\Pi}}'_{\Delta,**} = \nu \hat{\beta}_{1\wedge} \hat{\delta}_t \hat{\delta}'_t = 0_{G\wedge}$$

oder weiter

$$(9.57)\quad \hat{\beta}_{1\Delta}(\hat{\tilde{\Pi}}_{\Delta,**} \tilde{x}_{t**} \tilde{x}'_{t**} \hat{\tilde{\Pi}}'_{\Delta,**} - \nu \hat{\delta}_t \hat{\delta}'_t) = 0_{G\wedge}$$

Wir sehen, daß wir eine nicht triviale Schätzung $\hat{\beta}_{1\wedge}$ nur erhalten, wenn die in der Klammer stehende Matrix der Ordnung $G^{\wedge}$ singulär ist, d.h. wenn für ihre Determinante gilt

$$(9.58)\quad \left| \hat{\tilde{\Pi}}_{\wedge,**} \tilde{x}_{t**} \tilde{x}'_{t**} \hat{\tilde{\Pi}}'_{\wedge,**} - \nu \hat{\delta}_t \hat{\delta}'_t \right| = 0$$

Unter den Wurzeln der aus dieser Determinantengleichung entwickelten charakteristischen Gleichung wählen wir diejenige Wurzel $\hat{\nu}$ aus, die am kleinsten ist, setzen sie in (9.57) ein und können dann nach den gesuchten Schätzungen $\hat{\beta}_{1\wedge}$ auflösen. $\hat{\gamma}_{1*}$ können wir dann berechnen mittels der analog zu (9.45) mit (9.47) gewonnenen Gleichung

$$(9.59)\quad \hat{\gamma}_{1*} = -\hat{\beta}_{1\Delta} \hat{\tilde{\Pi}}_{\Delta*} = \hat{\beta}_{1\Delta} y_{t\Delta} x'_{t*} (x_{t*} x'_{t*})^{-1},$$

wenn wir in (9.45) $\beta_{1\Delta}$ durch $\hat{\beta}_{1\Delta}$ und γ_{1*} durch $\hat{\gamma}_{1*}$ ersetzen.

---

1) Wegen (9.47) und (9.48) ist nur noch $\beta_{1\wedge}$ variabel.

2) Dieses stellt den minimalen "Widerspruch" dar. Vgl. S. 84.

Der Vollständigkeit halber sollen diese Schätzungen nun noch nach der Methode der größten Dichte abgeleitet werden.[1)]

Wir nehmen an, daß die latenten Variablen der reduzierten Form δ_t zeitlich unabhängig und normal verteilt sind mit Mittelwert Null und Streuungs-Kovarianz-Matrix

$$(9.60) \qquad E(\delta_t \, \delta_t') = \Omega.$$

Ferner nehmen wir an, daß wir je T Beobachtungen aller gemeinsam abhängigen und vorherbestimmten Variablen zur Verfügung haben und die vorherbestimmten Variablen fest vorgegeben sind. Außerdem sei unsere Strukturgleichung (9.3) identifizierbar. Die Likelihoodfunktion der Gleichung (9.37) lautet dann[2)], da bei der reduzierten Form die Jacobische Matrix die Einheitsmatrix I_{G^Δ} der Ordnung G^Δ ist und entsprechend die Jacobische Determinante dem Skalar 1 gleich ist,

$$(9.61) \quad (y_{t_\Delta} | x_t) = (2\pi)^{\frac{-TG^\Delta}{2}} (\det \Omega^{-1})^{\frac{T}{2}} |\det I_{G^\Delta}|^T \cdot \exp\left(-\frac{1}{2} \sum_{t=1}^{T} \delta_t' \, \Omega^{-1} \, \delta_t\right)$$

Dies ist gleich

$$(9.62) \quad (y_{t_\Delta} | x_t) = (2\pi)^{\frac{-TG^\Delta}{2}} (\det \Omega^{-1})^{\frac{T}{2}} \cdot \exp\{-\frac{1}{2} \sum_{t=1}^{T} [(y_{t_\Delta} - \Pi_{\Delta^*} x_{t*} - \Pi_{\Delta,**} x_{t**})' \cdot \Omega^{-1} (y_{t_\Delta} - \Pi_{\Delta^*} x_{t*} - \Pi_{\Delta,**} x_{t**})]\}$$

---

1) Vgl. zum folgenden: Anderson u. Rubin (1949): a.a.O., S. 52-54.

2) Vgl. (8.14).

Die Likelihoodfunktion ist folgenden Restriktionen unterworfen

$$(9.16) \qquad 0_{1**} = \beta_{1\Delta}\, \Pi_{\Delta,**}$$

und der Normierung[1)]

$$(9.63) \qquad \beta_{1\Delta}\, \Phi\, \beta'_{1\Delta} = 1,$$

wobei Φ entweder die bekannte Konstante Ω oder eine bekannte Funktion der unbekannten Parameter von $\beta_{1\Delta}$ ist.

Wenn wir nun entsprechend die Likelihoodfunktion von (9.40), des reduzierte-Form-Untergleichungssystems mit der "neuen" Variablen $\tilde{x}_{t**}$, bilden, so können wir zum Zwecke der nachfolgenden Maximierung die mit Lagrange-Multiplikatoren versehenen Restriktionen (9.63) und (9.46) addieren.

Wenn wir dann den logarithmus naturalis der Likelihoodfunktion bilden und durch T dividieren, erhalten wir die logarithmische Likelihoodfunktion

$$(9.64) \quad L = -\frac{G\Delta}{2} \ln(2\pi) + \frac{1}{2}\ln \det \Omega^{-1} + \varphi(\beta_{1\Delta}\Phi\beta'_{1\Delta}-1)+\lambda'\tilde{\pi}'_{\Delta,**}\beta'_{1\Delta}$$

$$- \frac{1}{2T}\sum_{t=1}^{T}[(y_{t\Delta} - \tilde{\Pi}_{\Delta*}x_{t*} - \tilde{\Pi}_{\Delta,**}\tilde{x}_{t**})'$$

$$\Omega^{-1}(y_{t\Delta} - \tilde{\Pi}_{\Delta*}\, x_{t*} - \tilde{\Pi}_{\Delta,**}\, \tilde{x}_{t**})],$$

wobei φ ein Skalar und λ ein Spaltenvektor der Ordnung (K** x 1) sind.

Die Differentiation von (9.64) nach $\beta_{1\Delta}$ ergibt

$$(9.65) \qquad \frac{\partial L}{\partial \beta_{1\Delta}} = 2\varphi\, \beta_{1\Delta}\, \Phi + \lambda'\tilde{\Pi}'_{\Delta,**} = 0$$

Wenn wir (9.65) von rechts mit $\beta_{1\Delta}$ multiplizieren, erhalten wir

$$2\varphi\, \beta_{1\Delta} \Phi \beta'_{1\Delta} + \lambda' \tilde{\Pi}'_{\Delta,**}\, \beta'_{1\Delta} = 0 \quad . \tag{9.66}$$

Da $\tilde{\Pi}_{\Delta,**}\, \beta'_{1\Delta}$ wegen (9.46) ein aus G^{Δ} Nullen bestehender Spaltenvektor ist, muß wegen (9.63) der Lagrange-Multiplikator $\varphi = 0$ sein. Somit wird die Maximierung durch die Restriktion (9.63) nicht beeinflußt. Es steht uns also frei, von der Normierung (9.63) erst nach der Maximierung Gebrauch zu machen.

Der leichteren Rechenbarkeit halber können wir also folgende - das Maximum nicht beeinflussende - Normierung der Streuung von u_{1t}

$$\beta_{1\Delta}\, \Omega\, \beta'_{1\Delta} = 1 \tag{9.67}$$

verwenden.

Durch partielle Differentiation von (9.64) nach $\Omega, \tilde{\Pi}_{\Delta*}, \tilde{\Pi}_{\Delta,**}$ und $\beta_{1\Delta}$ sowie Gleichsetzen dieser Ableitungen mit Null erhalten wir[1)]

$$\begin{aligned} \hat{\Omega} = {} & y_{t\Delta}\, y'_{t\Delta} - y_{t\Delta}\, x'_{t*}\, \hat{\tilde{\Pi}}'_{\Delta*} - y_{t\Delta}\, \tilde{x}'_{t**}\, \hat{\tilde{\Pi}}'_{\Delta,**} \\ & - \hat{\tilde{\Pi}}_{\Delta*}\, x_{t*}\, y'_{t\Delta} - \hat{\tilde{\Pi}}_{\Delta,**}\, \tilde{x}_{t**}\, y'_{t\Delta} + \hat{\tilde{\Pi}}_{\Delta*}\, x_{t*}\, x'_{t*}\, \hat{\tilde{\Pi}}'_{\Delta*} \\ & + \hat{\tilde{\Pi}}_{\Delta,**}\, \tilde{x}_{t**}\, \tilde{x}'_{t**}\, \hat{\Pi}'_{\Delta,**} \end{aligned} \tag{9.68}$$

1) Vgl. Menges (1961): a.a.O., S. 225 und die im 2. und 3. Kapitel dieser Arbeit entwickelten Differentiationsregeln. Außerdem ist darauf zu achten, daß $\tilde{x}_{t**}$ orthogonal zu x_{t*} ist.

$$(9.69)\qquad \hat{\Omega}^{-1}(\hat{\tilde{\Pi}}_{\Delta *}\, x_{t*}\, x'_{t*} - y_{t_\Delta}\, x'_{t*}) = 0$$

$$(9.70)\qquad -\hat{\Omega}^{-1}(\hat{\tilde{\Pi}}_{\Delta,**}\, \tilde{x}_{t**}\, \tilde{x}'_{t**} - y_{t_\Delta}\, \tilde{x}'_{t**}) + \hat{\beta}'_{1_\Delta}\, \hat{\lambda}' = 0$$

$$(9.71)\qquad \hat{\lambda}'\, \hat{\tilde{\Pi}}'_{\Delta,**} = 0.$$

Wenn wir (9.69) nach $\hat{\tilde{\Pi}}_{\Delta *}$ auflösen, erhalten wir [Vgl. auch (9.47)]

$$(9.72)\qquad \hat{\tilde{\Pi}}_{\Delta *} = y_{t_\Delta}\, x'_{t*}(x_{t*}\, x'_{t*})^{-1}.$$

Die Auflösung von (9.70) nach $\hat{\tilde{\Pi}}_{\Delta,**}$ ergibt

$$(9.73)\qquad \hat{\tilde{\Pi}}_{\Delta,**} = \hat{\Omega}\, \hat{\beta}'_{1_\Delta}\, \hat{\lambda}'(\tilde{x}_{t**}\, \tilde{x}'_{t**})^{-1} + y_{t_\Delta}\tilde{x}'_{t**}(\tilde{x}_{t**}\tilde{x}'_{t**})^{-1}.$$

Durch Multiplikation von (9.73) mit $\hat{\beta}_{1_\Delta}$ von links und Auflösen nach $\hat{\lambda}'$ gelangen wir wegen (9.46) zu

$$(9.74)\qquad \hat{\lambda}' = -\hat{\beta}_{1_\Delta}\, y_{t_\Delta}\, \tilde{x}'_{t**},$$

was wir wieder in (9.73) einsetzen. Dann erhalten wir

$$(9.75)\qquad \hat{\tilde{\Pi}}_{\Delta,**} = y_{t_\Delta}\, \tilde{x}'_{t**}(\tilde{x}_{t**}\, \tilde{x}'_{t**})^{-1} - \hat{\Omega}\, \hat{\beta}'_{1_\Delta}\, \hat{\beta}_{1_\Delta}\, y_{t_\Delta}\, x'_{t**}$$
$$\cdot\, (\tilde{x}_{t**}\, \tilde{x}'_{t**})^{-1}$$
$$= (I_{G\Delta} - \hat{\Omega}\, \hat{\beta}'_{1_\Delta}\, \hat{\beta}_{1_\Delta})P_{\Delta,**},$$

wobei $P_{\Delta,**}$ steht für

$$(9.76) \qquad P_{\Delta,**} = y_{t\Delta}\,\tilde{x}'_{t**}(\tilde{x}_{t**}\,\tilde{x}'_{t**})^{-1}.$$

Mit Hilfe von (9.72), (9.75) und (9.50) gelangen wir von (9.68) zu

$$(9.77) \qquad \hat{\Omega} = \hat{\delta}_t\hat{\delta}'_t + \hat{\Omega}\,\hat{\beta}'_{1\Delta}\hat{\beta}_{1\Delta}P_{\Delta,**}\tilde{x}_{t**}\tilde{x}'_{t**}P'_{\Delta,**}\hat{\beta}'_{1\Delta}\hat{\beta}_{1\Delta}\hat{\Omega}.$$

Nun setzen wir zu Vereinfachung für das Skalarprodukt

$$(9.78) \qquad \hat{\beta}_{1\Delta}\,P_{\Delta,**}\,\tilde{x}_{t**}\,\tilde{x}'_{t**}\,P'_{\Delta,**}\,\hat{\beta}'_{1\Delta} = u.$$

Durch Multiplikation von (9.77) mit $\hat{\beta}'_{1\Delta}$ von rechts und bei Berücksichtigung von (9.78) und (9.67) erhält man

$$(9.79) \qquad \hat{\Omega}\,\hat{\beta}'_{1\Delta} = \hat{\delta}_t\,\hat{\delta}'_t\,\hat{\beta}'_{1\Delta} + \hat{\Omega}\,\hat{\beta}'_{1\Delta}\,u$$

und hieraus

$$(9.80) \qquad \hat{\Omega}\,\hat{\beta}'_{1\Delta} = \frac{1}{1-u}\,\hat{\delta}_t\,\hat{\delta}'_t\,\hat{\beta}'_{1\Delta}.$$

Aus (9.71) erhalten wir wegen (9.74), (9.75), (9.76) und (9.78) nunmehr

$$(9.81) \qquad \hat{\beta}_{1\Delta}\,P_{\Delta,**}\,\tilde{x}_{t**}\,\tilde{x}'_{t**}\,P'_{\Delta,**} - u\,\hat{\beta}_{1\Delta}\,\hat{\Omega} = 0.$$

Wenn wir (9.80) transponieren ($\hat{\Omega}$ ist symmetrisch) und in (9.81) einsetzen, erhalten wir

$$(9.82) \qquad \hat{\beta}_{1\Delta}(P_{\Delta,**}\,\tilde{x}_{t**}\,\tilde{x}'_{t**}\,P'_{\Delta,**} - v\,\hat{\delta}_t\,\hat{\delta}'_t) = 0.$$

wobei

$$(9.83) \qquad v = \frac{u}{1-u}.$$

(9.82) stimmt genau überein mit (9.57), wie man durch Vergleich von (9.76) mit (9.48) feststellen kann.

10. Die Methode der größten Dichte bei beschränkter Information nach Koopmans und Hood[1)]

Die Herausgeber der 14. Monographie der, nach ihrem Präsidenten Alfred Cowles benannten, Cowles Commission for Research in Economics, nämlich Tjalling C. Koopmans und William C. Hood, veröffentlichten 1953 in dieser Monographie eine neue Variante der Methode der größten Dichte bei beschränkter Information.

Im Gegensatz zu der vorangehenden Darstellung der Methode der größten Dichte bei beschränkter Information durch Anderson und Rubin gehen Koopmans und Hood von einem Teilsystem des Gleichungssystems statt von einer einzigen Gleichung aus. Dieses Teilsystem kann jedoch wie bei Anderson und Rubin auch nur aus einer einer einzigen Gleichung bestehen[2)].

Wie bei der von Anderson und Rubin dargestellten Methode werden bei der Methode der größten Dichte bei beschränkter Information nach Koopmans und Hood diejenigen a-priori-Restriktionen vernachlässigt, die die übrigen - nicht in diesem Teilsystem enthaltenen - Gleichungen betreffen. Es interessieren lediglich die a-priori-Informationen über die Variablen der Gleichungen des Teilsystems, insbesondere, welche Variablen des Teilsystems durch Null-Restriktionen betroffen sind.

Wir betrachten ein Gleichungssystem der Form

$$(10.1) \qquad Ax_t' = By_t' + \Gamma z_t' = u_t' \qquad (t=1,\ldots,T),$$

1) Koopmans u. Hood: a.a.O., S. 112-199, s. bes. S. 162-177 u. Appendix S. 185-199.

2) Koopmans u. Hood: a.a.O., S. 163.

wobei die Definitionen von Kapitel 8, Gleichungen (8.1) bis (8.4) gelten.

Dieses Gleichungssystem zerlegen wir nun in ein uns interessierendes Teilsystem I und in ein uns nicht interessierendes Restsystem II. Entsprechend zerlegen wir die Parametermatrix A in

$$(10.2) \qquad A = \begin{bmatrix} A_I \\ A_{II} \end{bmatrix}$$

und den Vektor der latenten Variablen u_t' in

$$(10.3) \qquad u_t' = \begin{bmatrix} u_{It}' \\ u_{IIt}' \end{bmatrix} .$$

Da für die Streuungs-Kovarianz-Matrix S der latenten Variablen gilt

$$(10.4) \qquad E(u_t' \, u_t) = S$$

lautet die Zerlegung von S

$$(10.5) \quad E \begin{bmatrix} u_{It}' u_{It} & u_{It}' u_{IIt} \\ u_{IIt}' u_{It} & u_{IIt}' u_{IIt} \end{bmatrix} = S = \begin{bmatrix} S_{I\,I} & S_{I\,II} \\ S_{II\,I} & S_{II\,II} \end{bmatrix} .$$

Wenn wir[1] also alle Informationen, die das Restsystem II betreffen, vernachlässigen wollen, so sind nur noch die Matrizen A_I und $S_{I\,I}$ für uns von Interesse. A_I läßt sich noch weiter zerlegen in

1) Vgl. im folgenden Koopmans u. Hood: a.a.O., Appendix E, S. 192-195.

$$(10.6) \qquad A_I = \begin{bmatrix} B_I & \Gamma_I \end{bmatrix} .$$

Dagegen werden die Matrizen A_{II}, $S_{I\,II}$ $(= S'_{II\,I})$ und $S_{II\,II}$, die für uns ohne Interesse sind, durch geeignete Lineartransformationen umgestaltet:

Unsere Lineartransformationsmatrix laute

$$(10.7) \qquad L = \begin{bmatrix} I_{I\,I} & 0_{I\,II} \\ L_{II\,I} & L_{II\,II} \end{bmatrix},$$

wobei $L_{II\,II}$ regulär sein soll.
Dann erhalten wir die transformierte Matrix S^+ aus S durch[1)2)]

$$(10.8) \quad S^+ = LSL' = \begin{bmatrix} I_{I\,I} & 0_{I\,II} \\ L_{II\,I} & L_{II\,II} \end{bmatrix} \begin{bmatrix} S_{I\,I} & S_{I\,II} \\ S_{II\,I} & S_{II\,II} \end{bmatrix} \begin{bmatrix} I_{I\,I} & L'_{II\,I} \\ 0_{II\,I} & L'_{II\,II} \end{bmatrix}$$

$$= \begin{bmatrix} S_{I\,I} & S_{I\,II} \\ L_{II\,I}S_{I\,I}+L_{II\,II}S_{II\,I} & L_{II\,I}S_{I\,II}+L_{II\,II}S_{II\,II} \end{bmatrix}$$

$$\cdot \begin{bmatrix} I_{I\,I} & L'_{II\,I} \\ 0_{II\,I} & L'_{II\,II} \end{bmatrix},$$

so daß

---

1) Vgl. Koopmans u. Hood: a.a.O., S. 190.

2) Das Symbol $^+$ soll verdeutlichen, daß es sich um eine Lineartransformation handelt.

$$(10.9) \qquad S^{+} = \begin{bmatrix} S^{+}_{I\ I} & S^{+}_{I\ II} \\ S^{+}_{II\ I} & S^{+}_{II\ II} \end{bmatrix}$$

$$= \begin{bmatrix} S_{I\ I} & S_{I\ I} L'_{II\ I} + S_{I\ II} L'_{II\ II} \\ L_{II\ I} S_{I\ I} + L_{II\ II} S_{II\ I} & S^{+}_{II\ II} \end{bmatrix},$$

wobei

$$(10.9a) \quad S^{+}_{II\ II} = L_{II\ I} S_{I\ I} L'_{II\ I} + L_{II\ II} S_{II\ I} L'_{II\ I} + L_{II\ I} S_{I\ II} L'_{II\ II} + L_{II\ II} S_{II\ II} L_{II\ II}$$

Die Transformation läßt also $S_{I\ I}$ unberührt, was mit (10.10) nochmals verdeutlicht werden soll

$$(10.10) \quad S^{+} = L \begin{bmatrix} S_{I\ I} & S'_{II\ I} \\ S_{II\ I} & S_{II\ II} \end{bmatrix} L' = \begin{bmatrix} S_{I\ I} & S^{+'}_{II\ I} \\ S^{+}_{II\ I} & S^{+}_{II\ II} \end{bmatrix},$$

wobei $S^{+'}_{II\ I} = S^{+}_{I\ II}$, wie man leicht mit (10.9) zeigen kann.

Um nun zu erreichen, daß $S^{+}_{I\ II} = 0_{I\ II}$, wählen wir für das bislang nicht genauer spezifizierte $L'_{II\ I}$ den Wert

$$(10.11) \quad L'_{II\ I} = - S^{-1}_{I\ I} S_{I\ II} L'_{II\ II} .$$

Durch Einsetzen von (10.11) in (10.9) erhalten wir dann[1]

$$(10.12) \quad S^{+} = \begin{bmatrix} S_{I\ I} & 0_{I\ II} \\ 0_{II\ I} & S^{+}_{II\ II} \end{bmatrix}$$

---

1) Wir haben nun zwei Untersysteme, bei denen die Streuungen eines Untersystems unabhängig sind von den Streuungen des anderen Untersystems.

wobei $S^{+}_{II\ I} = 0_{II\ I}$ ebenfalls wegen (10.11), was sich durch Einsetzen der transponierten Matrix $L_{II\ I}$ von $L'_{II\ I}$ in (10.9) leicht zeigen läßt, wobei wegen (10.11) gilt

$$L_{II\ I} = -L_{II\ II}\, S_{II\ I}\, S^{-1}_{I\ I}. \tag{10.13}$$

Mit Hilfe von (10.13) können wir außerdem einen kürzeren Ausdruck für $S^{+}_{II\ II}$ erhalten, wie (10.14) zeigen soll[1]

$$\begin{aligned} S^{+}_{II\ II} &= -L_{II\ II}\, S_{II\ I}\, L'_{II\ I} + L_{II\ II}\, S_{II\ I}\, L'_{II\ I} \\ &\quad -L_{II\ II}\, S_{II\ I}\, S^{-1}_{I\ I}\, S_{I\ II}\, L'_{II\ II} \\ &\quad +L_{II\ II}\, S_{II\ II}\, L'_{II\ II} \\ &= L_{II\ II}(S_{II\ II} - S_{II\ I}\, S^{-1}_{I\ I}\, S_{I\ II})L'_{II\ II}, \end{aligned} \tag{10.14}$$

weil die beiden ersten Ausdrücke sich gegeneinander aufheben. Nachdem wir die Lineartransformationsmatrix L bisher nur durch (10.7) und (10.11) restringiert haben, verwenden wir nun den letzten Freiheitsgrad zur Bestimmung des Wertes von $L_{II\ II}$. Wir wählen den Wert für $L_{II\ II}$ derart, daß

$$\begin{aligned} S^{+}_{II\ II} &= L_{II\ II}(S_{II\ II} - S_{II\ I}\, S^{-1}_{I\ I}\, S_{I\ II})L'_{II\ II} \\ &= I_{II\ II} \end{aligned} \tag{10.15}$$

Damit erhalten wir für S^{+} die Gestalt

$$S^{+} = LSL' = \begin{bmatrix} S_{I\ I} & 0_{I\ II} \\ 0_{II\ I} & I_{II\ II} \end{bmatrix} \tag{10.16}$$

1) Vgl. mit (10.9).

Das Ergebnis dieser Lineartransformation können wir in die logarithmische Likelihoodfunktion (10.17) einsetzen[1)]

$$(10.17)\quad \mathfrak{L} = -\frac{G}{2}\ln(2\pi) - \frac{1}{2}\ln(\det S) + \ln|\det B| - \frac{1}{2}\,\mathrm{Sp}(S^{-1} A M_{xx} A'),$$

wenn wir berücksichtigen, daß äquivalente Strukturen, die sich durch solche Lineartransformationen herleiten lassen, dieselbe logarithmische Likelihoodfunktion aufweisen, nämlich[2)]

$$(10.18)\quad \mathfrak{L} = \mathfrak{L}^{+} = -\frac{G}{2}\ln(2\pi) - \frac{1}{2}\ln(\det LSL') + \ln|\det LB| - \frac{1}{2}\,\mathrm{Sp}[(LSL')^{-1}(LA)M_{xx}(A'L')]$$

$$= -\frac{G}{2}\ln(2\pi) - \frac{1}{2}\ln(\det S_{I\,I}) + \ln|\det LB| - \frac{1}{2}\,\mathrm{Sp}\left\{\begin{bmatrix} S_{I\,I}^{-1} & 0_{I\,II} \\ 0_{II\,I} & I_{II\,II} \end{bmatrix}\begin{bmatrix} A_I \\ A_{II}^{+} \end{bmatrix} M_{xx}\begin{bmatrix} A_I' & A_{II}^{+\prime} \end{bmatrix}\right\},$$

wobei

$$(10.19)\quad LA = \begin{bmatrix} I_{I\,I} & 0_{I\,II} \\ L_{II\,I} & L_{II\,II} \end{bmatrix}\begin{bmatrix} A_I \\ A_{II} \end{bmatrix} = \begin{bmatrix} A_I \\ A_{II}^{+} \end{bmatrix}$$

mit

$$(10.20)\quad A_{II}^{+} = L_{II\,I}\,A_I + L_{II\,II}\,A_{II}$$

---

1) Vgl. (8.17).

2) Vgl. zur Identität von $\mathfrak{L}^{+}$ mit $\mathfrak{L}$ die Gleichung (8.20).

und

$$(10.21)\quad LB = \begin{bmatrix} I_{I\,I} & 0_{I\,II} \\ L_{II\,I} & L_{II\,II} \end{bmatrix} \begin{bmatrix} B_I \\ B_{II} \end{bmatrix} = \begin{bmatrix} B_I \\ L_{II\,I}B_I + L_{II\,II}B_{II} \end{bmatrix} .$$

Die Ausmultiplikation von (10.18) ergibt

$$(10.22)\quad \varrho = -\frac{G}{2}\ln(2\pi) - \frac{1}{2}\ln(\det S_{I\,I}) + \ln|\det LB| - \frac{1}{2}\,Sp\begin{bmatrix} S_{I\,I}^{-1} A_I M_{xx} A'_I & S_{I\,I}^{-1} A_I M_{xx} A_{II}^{+} \\ A_{II}^{+} M_{xx} A'_I & A_{II}^{+} M_{xx} A_{II}^{+'} \end{bmatrix} .$$

Wie man sich leicht verdeutlichen kann, ist die Spur des letzten Ausdrucks, also die Summe der Hauptdiagonalelemente der Gesamtmatrix, gleich der Summe der Spuren jener Teilmatrizen, die Elemente auf jener Hauptdiagonalen besitzen. Die Spur der Gesamtmatrix ist demnach die Summe der Elemente auf der Hauptdiagonalen der Untermatrizen $S_{I\,I}^{-1} A_I M_{xx} A'_I$ und $A_{II}^{+} M_{xx} A_{II}^{+'}$.

Wir erhalten

$$(10.23)\quad \varrho = -\frac{G}{2}\ln(2\pi) - \frac{1}{2}\ln(\det S_{I\,I}) + \ln|\det LB| - \frac{1}{2}\left[Sp(S_{I\,I}^{-1} A_I M_{xx} A'_I) + Sp(A_{II}^{+} M_{xx} A_{II}^{+'})\right]$$

und nach Auflösen der letzten Klammer

$$(10.24)\quad \varrho = -\frac{G}{2}\ln(2\pi) - \frac{1}{2}\ln(\det S_{I\,I}) + \ln|\det LB| - \frac{1}{2}\,Sp(S_{I\,I}^{-1} A_I M_{xx} A'_I) - \frac{1}{2}\,Sp(A_{II}^{+} M_{xx} A_{II}^{+'})$$

In (10.24) enthalten der dritte und der fünfte Ausdruck auf der rechten Seite der Gleichung noch Informationen

über das Restsystem II. Durch schrittweise Differentiation von (10.24) soll im folgenden der Einfluß des Restsystems II beseitigt werden.

Zunächst soll (10.24) bezüglich A_{II}^{+} differenziert werden[1]. Da nur der dritte und der fünfte Ausdruck auf der rechten Seite von (10.24) Elemente von A_{II}^{+} aufweisen, erhalten wir[2]

$$(10.25)\quad \frac{\partial\, \Omega}{\partial\, A_{II}^{+}} = \frac{\partial \ln|\det LB|}{\partial\, A_{II}^{+}} = \frac{\partial\, \frac{1}{2}\, Sp(A_{II}^{+}\, M_{xx}\, A_{II}^{+'})}{\partial\, A_{II}^{+}}$$

$$= Sp\,(B'L')^{-1}\, \frac{\partial\, B'L'}{\partial\, A_{II}^{+}} - Sp(A_{II}^{+} M_{xx} I_{G^{\Delta\Delta},G^{\Delta\Delta}+K^{**}})$$

$$= Sp\,(B'L')^{-1}\, \frac{\partial\, B'L'}{\partial\, A_{II}^{+}} - G^{\Delta\Delta}(A_{II}^{+}\, M_{xx}) = 0,$$

weil in der Einheitsmatrix $I_{G^{\Delta\Delta}[(G^{\Delta\Delta}+K^{**})\times(G^{\Delta\Delta}+K^{**})]G^{\Delta\Delta}}$ auf der Hauptdiagonalen $G^{\Delta\Delta}$ Einheitsmatrizen der Ordnung $I_{(G^{\Delta\Delta}+K^{**})}$ stehen.

Um den Ausdruck $\frac{\partial\, L\, B}{\partial\, A_{II}^{+}}$ besser zu verstehen, betrachten wir das Matrizenprodukt LA, das LB und A_{II}^{+} enthält:

$$(10.26)\quad LA = \begin{bmatrix} I_{I\ I} & 0_{I\ II} \\ 0_{II\ I} & L_{II\ II} \end{bmatrix} \begin{bmatrix} B_{I} & \Gamma_{I} \\ B_{II} & \Gamma_{II} \end{bmatrix} = \begin{bmatrix} B_{I}^{+} & \Gamma_{I}^{+} \\ B_{II}^{+} & \Gamma_{II}^{+} \end{bmatrix}$$

---

1) Der besseren Übersicht halber soll hier nur die Differentiation der Matrizen en bloc gemäß den Differentiationsregeln für Matrizen vorgenommen werden. Elementweise Differentiation wird von Koopmans u. Hood, a.a.O., im Anhang E angedeutet.

2) Vgl. die Differentiationsregeln in: Menges (1961): a.a.O., S. 225 und die Differentiation mit Matrizen im 3. Kapitel dieser Arbeit.

Da LB = $\begin{bmatrix} B_I^+ \\ B_{II}^+ \end{bmatrix}$ und $A_{II}^+ = \begin{bmatrix} B_{II}^+ & \Gamma_{II}^+ \end{bmatrix}$ nur die Matrix B_{II}^+ gemeinsam haben, ist das Ergebnis der Differentiation $\frac{\partial B'L'}{\partial A_{II}^+}$ eine Matrix vom Typ

$$(10.27) \qquad \begin{bmatrix} 0 & 0 \\ I_{G^{\Delta\Delta},G^{\Delta\Delta}} & 0 \end{bmatrix} ,$$

wobei $I_{G^{\Delta\Delta},G^{\Delta\Delta}}$ das Ergebnis der Differentiation $\frac{\partial B_{II}^{+'}}{\partial B_{II}^+}$ ist. Da $-A_{II}^+ M_{xx}$ nur eine Matrix der Ordnung $[G^{\Delta\Delta}{}_x(G^{\Delta\Delta} + K^{**})]$ ist, ist für die Subtraktion in (10.25) nur die Untermatrix $[I \quad 0]$ von $\begin{bmatrix} 0 & 0 \\ I & 0 \end{bmatrix}$ beachtlich und wir erhalten für das Untersystem II

$$(10.28) \qquad \frac{\partial \varrho}{\partial A_{II}^+} = (B'L')_{II}^{-1} [I_{G^{\Delta\Delta}} \quad 0] - A_{II}^+ M_{xx} = 0 ,$$

weil die Spur von $G^{\Delta\Delta}$ identischen Untermatrizen auf der Hauptdiagonalen gleich dem $G^{\Delta\Delta}$-fachen dieser Untermatrizen ist und wir nun durch $G^{\Delta\Delta}$ dividieren können. Hieraus folgt

$$(10.29) \qquad A_{II}^+ M_{xx} = (B'L')_{II}^{-1} [I_{G^{\Delta\Delta}} \quad 0]$$

Wir können jetzt an Stelle des fünften Ausdrucks auf der rechten Seite der Gleichung (10.24) setzen

$$(10.30)\quad -\tfrac{1}{2}\,\mathrm{Sp}(A_{II}^{+}M_{xx}A_{II}^{+'}) = -\tfrac{1}{2}\,\mathrm{Sp}\left\{(B'L')_{II}^{-1}[I_{G\Delta\Delta}\ 0]\begin{bmatrix}B_{II}^{+'}\\ \Gamma_{II}^{+'}\end{bmatrix}\right\}$$

$$= -\tfrac{1}{2}\,\mathrm{Sp}\left[(B'L')_{II}^{-1}\,B_{II}^{+'}\right]$$

$$= -\tfrac{1}{2}\,\mathrm{Sp}\left[(B_{II}^{+'})^{-1}\,B_{II}^{+'}\right]$$

$$= -\tfrac{1}{2}\,\mathrm{Sp}\,(I_{G\Delta\Delta})$$

$$= -\frac{G^{\Delta\Delta}}{2}\quad ,$$

wobei $G^{\Delta\Delta}$ die Zahl der endogenen Variablen von System II ist, also eine Konstante, die für die Bestimmung des Maximums der logarithmierten Likelihoodfunktion unwesentlich ist.

Es folgt des weiteren aus

$$(10.31)\quad A_{II}^{+}\,M_{xx} = [B_{II}^{+}\quad \Gamma_{II}^{+}]\begin{bmatrix}M_{yy} & M_{yz}\\ M_{zy} & M_{zz}\end{bmatrix}$$

und (10.29), daß

$$(10.32)\quad (B'L')_{II}^{-1}\,[I_{G\Delta\Delta}\ 0] = [B_{II}^{+}M_{yy}+\Gamma_{II}^{+}M_{zy}\quad B_{II}^{+}M_{yz}+\Gamma_{II}^{+}M_{zz}]$$

Dann ist

$$(10.33)\quad (B'L')_{II}^{-1} = B_{II}^{+}\,M_{yy} + \Gamma_{II}^{+}\,M_{zy}$$

und

$$(10.34)\quad 0_{G\Delta\Delta} = B_{II}^{+}\,M_{yz} + \Gamma_{II}^{+}\,M_{zz}$$

Den aus (10.34) berechneten Wert für

$$\Gamma^{+}_{II} = -B^{+}_{II}\, M_{yz}\, M^{-1}_{zz} \tag{10.35}$$

setzen wir in (10.33) ein und erhalten

$$(B'L')^{-1}_{II} = B^{+}_{II}(M_{yy} - M_{yz}\, M^{-1}_{zz}\, M_{zy}) \tag{10.36}$$

$$= B^{+}_{II}\, W \ ,$$

wobei also

$$W = (M_{yy} - M_{yz}\, M^{-1}_{zz}\, M_{zy}). \tag{10.37}$$

Wir verwenden nun die Determinantenregel[1)]

$$\det (x\ y) = \det x \cdot \det y \tag{10.38}$$

und erhalten für

$$\ln \det (B^{+}\, W\, B^{+'}) = \ln \det B^{+} + \ln \det W + \ln \det B^{+'} \tag{10.39}$$

oder, umgestellt, erhalten wir für den dritten Ausdruck auf der rechten Seite von Gleichung (10.24)

$$\ln|\det B^{+}| = \tfrac{1}{2} \ln \det(B^{+}\, W\, B^{+'}) - \tfrac{1}{2} \ln \det W \tag{10.40}$$

Wenn wir (10.36) verwenden, erhalten wir für

---

1) Vgl. Zurmühl: a.a.O., S. 21.

$$(10.41)\quad B^{+}WB^{+'} = \begin{bmatrix} B_I^{+} \; W \; B^{+'} \\ B_{II}^{+} \; W \; B^{+'} \end{bmatrix} = \begin{bmatrix} B_I^{+} \quad W \quad B^{+'} \\ (B'L')_{II}^{-1} \; B^{+'} \end{bmatrix} = \begin{bmatrix} B_I^{+} \; W \; B^{+'} \\ (B_{II}^{+'})^{-1} B^{+'} \end{bmatrix}$$

$$= \begin{bmatrix} B_I^{+} \; W \; B_I^{+'} & B_I^{+} \; W \; B_{II}^{+'} \\ (B_{II}^{+'})^{-1} \; B_I^{+'} & (B_{II}^{+'})^{-1} \; B_{II}^{+'} \end{bmatrix}$$

$$= \begin{bmatrix} B_I^{+} \; W \; B_I^{+'} & B_I^{+} \; W \; B_{II}^{+'} \\ 0_{II\ I} & I_{II\ II} \end{bmatrix} \quad 1)$$

Dann ist

$$(10.42)\quad \ln \det (B^{+} \; W \; B^{+'}) = \ln \det (B_I^{+} \; W \; B_I^{+'}) = \ln \det (B_I \; W \; B_I')$$

wegen $B_I^{+} = B_I$ [Vgl. (10.21)].

Es folgt aus (10.42) und (10.40)

$$(10.43)\quad \ln|\det B^{+}| = \tfrac{1}{2} \ln \det(B_I \; W \; B_I') - \tfrac{1}{2} \ln \det W$$

Dann lautet die neue logarithmische Likelihoodfunktion wegen (10.24), (10.30) und (10.43)

---

1) Koopmans u. Hood (a.a.O., S. 195) zeigen nicht die Entwicklung von $(B_{II}^{+})^{-1} B_I^{+}$ zu $0_{II\ I}$, die für die weiteren Rechnungen mit der Determinante relevant ist. Obwohl der Verfasser dieser Arbeit alternative Definitionsgleichungen heranzog, konnte dieses Ergebnis nicht bestätigt werden. Stimmt dieses Ergebnis jedoch, so muß auch gelten $B_I^{+} \; W \; B_{II}^{+} = 0_{I\ II}$, da W symmetrisch ist. Für die Determinante ist letzteres allerdings unbeachtlich.

$$(10.44)\quad \varrho(A_I, S_{I\,I}) = -\frac{G}{2}\ln(2\pi) - \frac{G^{\Delta\Delta}}{2} - \frac{1}{2}\ln(\det S_{I\,I}) + \frac{1}{2}\ln\det(B_I\, W\, B_I') - \frac{1}{2}\ln\det W - \frac{1}{2}\mathrm{Sp}(S_{I\,I}^{-1}\, A_I\, M_{xx}\, A_I') .$$

Durch weiteres schrittweises Maximieren von (10.44) bezüglich $S_{I\,I}$ gelangen wir zu einer konzentrierten logarithmischen Likelihoodfunktion, die nur noch nach den Elementen von A_I maximiert werden muß.

Wir bilden also[1)]

$$(10.45)\quad \frac{\partial\,\varrho(A_I, S_{I\,I})}{\partial\, S_{I\,I}} = -\frac{1}{2}\frac{\partial\,\ln(\det S_{I\,I})}{\partial\, S_{I\,I}} - \frac{\partial\,\mathrm{Sp}(S_{I\,I}^{-1} A_I M_{xx} A_I')}{\partial\, S_{I\,I}}$$

$$= -\frac{1}{2}\mathrm{Sp}\left(S_{I\,I}^{-1}\frac{\partial\, S_{I\,I}}{\partial\, S_{I\,I}}\right) - \frac{1}{2}\mathrm{Sp}\left(\frac{\partial\, S_{I\,I}^{-1}}{\partial\, S_{I\,I}} A_I\, M_{xx}\, A_I'\right)$$

$$= -\frac{1}{2} S_{I\,I}^{-1} + \frac{1}{2}\left(S_{I\,I}^{-1}\, S_{I\,I}^{-1}\, A_I\, M_{xx}\, A_I'\right) = 0_{G\Delta} ,$$

Für

$$(10.46)\qquad S_{I\,I} = A_I\, M_{xx}\, A_I'$$

ist dann (10.45) gewährleistet, denn es gilt durch Einsetzen von (10.46) in (10.45)

$$(10.47)\qquad \frac{\partial\,\varrho\,(A_I,\, S_{I\,I})}{\partial\, S_{I\,I}} = -\frac{1}{2} S_{I\,I}^{-1} + \frac{1}{2} S_{I\,I}^{-1} = 0_{G\Delta}.$$

---

1) Es gelten die bereits bekannten Differentiationsregeln, wobei zu beachten ist, daß $S_{I\,I}$ symmetrisch ist.

Durch Einsetzen von (10.46) in (10.44) erhalten wir dann die auf A_I konzentrierte logarithmische Likelihoodfunktion

$$(10.48) \quad \wp(A_I) = -\frac{G}{2}\ln(2\pi) - \frac{G}{2} - \frac{1}{2}\ln\det(A_I M_{xx} A_I') + \frac{1}{2}\ln\det(B_I W B_I') - \frac{1}{2}\ln\det W,$$

weil die Spur des letzten Ausdrucks auf der rechten Seite von (10.44) gleich $G^{\wedge}$ ist.

"Die durch Maximierung dieser Funktion unter Berücksichtigung der a priori Restriktionen von A_I erhaltenen Schätzungen $\hat{A}_I$ von A_I sind limited-information-maximum-likelihood-Schätzungen, die jegliche Kenntnis der Restriktionen der Elemente von A_{II} außer acht lassen. Ihre Ableitung setzt jedoch voraus, daß das gesamte System linear ist, und, wegen der Art und Weise, in der M und W in (6.3)[1] erscheinen, daß es für ihre Berechnung notwendig ist, die vorherbestimmten Variablen zu kennen, die in den Gleichungen des gesamten Systems erscheinen."[2]

Koopmans und Hood[3] weisen darauf hin, daß die Kenntnis der endogenen Variablen des Restsystems II nicht notwendig ist, da diese nur in dem Ausdruck $-\frac{1}{2}\det W$ vorkommen und dieser Ausdruck, da er nicht A_I enthält, die Lage des Maximums (d.h. den Abszissenwert) nicht beeinflußt.[4]

---

1) (6.3) ist unsere Gleichung (10.48).

2) Koopmans und Hood, a.a.O., S. 165. Übersetzung des Verfassers. Für die Schätzungen von A_I wurde ein Dach verwendet.

3) Koopmans und Hood, a.a.O., S. 165, Fußnote 74.

4) Durch Ausmultiplizieren kann man zeigen, daß in $\frac{1}{2}\ln\det(B_I W B_I')$ der Teil von W, der das Restsystem II betrifft, wegfällt.

Wir wollen jetzt noch die Methode der größten Dichte bei beschränkter Information nach Koopmans und Hood für den Spezialfall einer einzigen Gleichung betrachten[1].

Wenn wir berücksichtigen, daß bei der Maximierung von (10.48) nur diejenigen Ausdrücke von Interesse sind, die die unbekannten Parameter (A_I) enthalten, so können wir statt (10.48) schreiben

$$(10.49) \qquad \varphi(A_I) = \varkappa - \frac{1}{2} \ln \det(A_I M_{xx} A_I') + \frac{1}{2} \ln \det(B_I W B_I')$$

wobei $\varkappa$ stellvertretend für die bei der Maximierung bezüglich A_I nicht relevanten Ausdrücke stehen soll[2].

Betrachten wir nun eine einzige Gleichung, nämlich die erste Gleichung des Gleichungssystems (10.1), welche folgendes Aussehen habe

$$(10.50) \qquad \alpha_1 x_t' + \beta_1 y_t' + \gamma_1 z_t' = u_{1t}' \qquad (t=1,\ldots,T),$$

wobei α_1, β_1 bzw. γ_1 jeweils die erste Zeile von A, B bzw. F sind.

Wenn wir die a-priori-(Null-)Restriktionen von α_1 und somit von β_1 und γ_1 berücksichtigen, so erhalten wir

$$(10.51) \qquad \alpha_{1(\Delta,*)} x_{t(\Delta,*)}' = \beta_{1\Delta} y_{t\Delta}' + \gamma_{1*} z_{t*}' = u_{1t}' \qquad (t=1,\ldots,T)$$ [3]

---

1) Koopmans u. Hood: a.a.O., S. 166-170.

2) Da insofern $\varkappa$ eine Konstante ist, entfällt sie beim Maximieren.

3) Im folgenden wird (t=1,...,T) nicht mitgeführt; gleichwohl soll der mitgeführte Index t die Möglichkeit der Zusammenfassung der Beobachtungswerte zu Vektoren und Matrizen vergegenwärtigen.

Dann ist im letzten Ausdruck auf der rechten Seite von (10.49) statt $B_I\, W\, B_I'$ nur noch die Untermatrix[1] $\beta_{1\Delta}\, W_{\Delta,\Delta}\, \beta_{1\Delta}'$ relevant, da uns hier nur die endogenen Variablen interessieren, die in (10.51) enthalten sind. Wir entnehmen also der Matrix

$$(10.52)\quad B_I\, W\, B_I' = \begin{bmatrix} \beta_{1\Delta}\, W_{\Delta,\Delta}\, \beta_{1\Delta}' & \beta_{1\Delta}\, W_{\Delta,\Delta\Delta}\, \beta_{1\Delta\Delta}' \\ \beta_{1\Delta\Delta}\, W_{\Delta\Delta,\Delta}\, \beta_{1\Delta}' & \beta_{1\Delta\Delta}\, W_{\Delta\Delta,\Delta\Delta}\, \beta_{1\Delta\Delta}' \end{bmatrix}$$

nur den linken oberen Ausdruck, da für (10.51) die Restriktion

$$(10.53)\qquad \beta_{1\Delta\Delta} = 0_{1\Delta\Delta}$$

gilt. Analog zu (10.37) ist $W_{\Delta\Delta}$ definiert durch

$$(10.54)\qquad W_{\Delta,\Delta} = (M_{y\Delta y\Delta} - M_{y\Delta z}\, M_{zz}^{-1}\, M_{zy\Delta})$$

Statt $A_I\, M_{xx}\, A_I'$ im zweiten Ausdruck auf der rechten Seite von (10.49) erhalten wir bei einer einzigen Gleichung vom Variablen ausschließenden Typ (10.51) nur

$$\frac{1}{T}\sum_{t=1}^{T} (\alpha_{1(\Delta,*)}\, x_{t(\Delta,*)}'\, x_{t(\Delta,*)}\, \alpha_{1(\Delta,*)}')$$

$$= \alpha_{1(\Delta,*)}\, M_{x(\Delta,*)}\, x_{(\Delta,*)}\, \alpha_{1(\Delta,*)}'$$

Dann können wir entwickeln

---

1) Diese Untermatrix ist ein Skalar.

$$(10.55)\quad \alpha_{1(\Delta,*)}\, M_{x(\Delta,*)}\, x_{(\Delta,*)}\, \alpha'_{1(\Delta,*)}$$

$$= \frac{1}{T}\sum_{t=1}^{T}[(\beta_{1\Delta}y'_{t\Delta} + \gamma_{1*}z'_{t*})(y_{t\Delta}\beta'_{1\Delta} + z_{t*}\gamma'_{1*})]$$

$$= \frac{1}{T}\sum_{t=1}^{T}(\beta_{1\Delta}y'_{t\Delta}y_{t\Delta}\beta'_{1\Delta} + \beta_{1\Delta}y'_{t\Delta}z_{t*}\gamma'_{1*} + \gamma_{1*}z_{t*}y_{t\Delta}\beta'_{1\Delta}$$

$$+ \gamma_{1*}z'_{t*}z_{t*}\gamma'_{1*}).$$

Setzen wir nun die Kleinst-Quadrat-Schätzung für γ_{1*}, die wir aus (10.51) herleiten können[1],

$$(10.56)\qquad \hat{\gamma}_{1*} = -\hat{\beta}_{1\Delta}y'_{t\Delta}z_{t*}(z'_{t*}z_{t*})^{-1}$$

in (10.55) ein, so gelangen wir zu

$$(10.57)\quad \alpha_{1(\Delta,*)}\, M_{x(\Delta,*)x(\Delta,*)}\, \alpha'_{1(\Delta,*)}$$

$$= \frac{1}{T}\sum_{t=1}^{T}(\hat{\beta}_{1\Delta}y'_{t\Delta}y_{t\Delta}\hat{\beta}'_{1\Delta} - \hat{\beta}_{1\Delta}y'_{t\Delta}z_{t*}(z'_{t*}z_{t*})^{-1}z'_{t*}y_{t\Delta}\hat{\beta}'_{1\Delta})$$

$$= \hat{\beta}_{1\Delta}\, W^{*}_{\Delta,\Delta}\, \hat{\beta}'_{1\Delta},$$

wobei

$$(10.58)\qquad W^{*}_{\Delta,\Delta} = (M_{y\Delta y\Delta} - M_{y\Delta z_*}\, M^{-1}_{z_*z_*}\, M_{z_*y\Delta})$$

Setzen wir nun die bei einer einzelnen Gleichung vom Typ (10.51) relevanten Ausdrücke in (10.49) ein, so erhalten wir

---

1) $\hat{\beta}_{1\Delta}y'_{t\Delta}$ kann hierbei als eine Variable y_t betrachtet werden. Vgl. auch das nur durch die Wahl eines anderen Vorzeichens abweichende γ_{1*} von Gleichung (11.16).

$$(10.59)\quad \varrho\,[\alpha_{1(\Delta,*)}] = \varkappa - \frac{1}{2}\ln\det(\hat{\beta}_{1\Delta} W_{\Delta,\Delta}^{*} \hat{\beta}'_{1\Delta}) + \frac{1}{2}\ln\det(\hat{\beta}_{1\Delta} W_{\Delta,\Delta} \hat{\beta}'_{1\Delta})$$

$$= \varkappa - \frac{1}{2}\ln(\hat{\beta}_{1\Delta} W_{\Delta,\Delta}^{*} \hat{\beta}'_{1\Delta}) + \frac{1}{2}\ln(\hat{\beta}_{1\Delta} W_{\Delta,\Delta} \hat{\beta}'_{1\Delta})$$

weil die Klammerausdrücke Skalare sind und die Determinante eines Skalars dem Skalar gleich ist.

Aus (10.59) erhalten wir

$$(10.60)\quad \varrho\,[\alpha_{1(\Delta,*)}] = \varkappa - \frac{1}{2}\ln(\hat{\beta}_{1\Delta} W_{\Delta,\Delta}^{*} \hat{\beta}'_{1\Delta} - \hat{\beta}_{1\Delta} W_{\Delta,\Delta} \hat{\beta}'_{1\Delta})$$

Wir sehen, daß (10.60) maximiert wird, wenn

$$(10.61)\quad \ln(\hat{\beta}_{1\Delta} W^{*}_{\Delta,\Delta} \hat{\beta}'_{1\Delta} - \hat{\beta}_{1\Delta} W_{\Delta,\Delta} \hat{\beta}'_{1\Delta}) = \frac{\hat{\beta}_{1\Delta} W^{*}_{\Delta,\Delta} \hat{\beta}'_{1\Delta}}{\hat{\beta}_{1\Delta} W_{\Delta,\Delta} \hat{\beta}'_{1\Delta}}$$

minimiert wird.

11. Das Prinzip des minimalen Streuungsverhältnisses[1)]

Wir betrachten wieder eine einzige Gleichung, nämlich die erste Gleichung eines vollständigen Gleichungssystems, welche folgendes Aussehen habe[2)]

$$(11.1) \qquad \beta_1 y_t' - \gamma_1 z_t' = u_{1t}' \qquad (t=1,\ldots,T)^{3)}.$$

Wenn wir die Nullrestriktionen (a-priori-Restriktionen) berücksichtigen, erhalten wir

$$(11.2) \qquad \beta_{1\Delta} y_{t\Delta}' - \gamma_{1*} z_{t*}' = u_{1t}'.$$

Wir bilden nun eine neue Variable für das erste Skalarprodukt auf der linken Seite von (11.2),

$$(11.3) \qquad \tilde{y}_t' = \beta_{1\Delta} y_{t\Delta}',$$

die wir in Abhängigkeit von z_{t*}, den in Gleichung (11.2) enthaltenen vorherbestimmten Variablen, schätzen.

Die Kleinst-Quadrat-Schätzung von γ_{1*} lautet

$$(11.4) \qquad \hat{\gamma}_{1*} = \tilde{y}_t' \, z_{t*} (z_{t*}' \, z_{t*})^{-1}.$$

Die Summe der Quadrate der Residuen $\hat{u}_{1t}$ hat folgendes Aussehen

---

1) Koopmans u. Hood: a.a.O., S. 166-170. Vgl. auch: Menges, G. (1961): a.a.O., S. 106-109 und Johnston: a.a.O., S. 254-258.

2) Vgl. wegen der Definitionen: (8.2) und (10.50). Aus Gründen der besseren Darstellbarkeit wurde das Vorzeichen der Parameter der vorherbestimmten Variablen negativ gewählt.

3) Im folgenden wird (t=1,...,T) nicht mitgeführt; gleichwohl soll der mitgeführte Index t die Möglichkeit der Zusammenfassung der Beobachtungswerte zu Vektoren und Matrizen vergegenwärtigen.

$$(11.5)\quad \sum_{t=1}^{T} \hat{u}_{1t}' \hat{u}_{1t} = \sum_{t=1}^{T} (\tilde{y}_t' - \hat{\gamma}_{1*} z_{t*}')(\tilde{y}_t - z_{t*} \hat{\gamma}_{1*}')$$

$$= \sum_{t=1}^{T} (\tilde{y}_t' \tilde{y}_t - \tilde{y}_t' z_{t*} \hat{\gamma}_{1*}' - \hat{\gamma}_{1*} z_{t*}' \tilde{y}_t + \hat{\gamma}_{1*} z_{t*}' z_{t*} \hat{\gamma}_{1*}')$$

$$= \sum_{t=1}^{T} (\tilde{y}_t' \tilde{y}_t - \hat{\gamma}_{1*} z_{t*}' z_{t*} \hat{\gamma}_{1*}') \ ^{1)}.$$

Hieraus erhalten wir wegen (11.3)

$$(11.6)\quad \sum_{t=1}^{T} \hat{u}_{1t}' \hat{u}_{1t} = \sum_{t=1}^{T} (\hat{\beta}_{1\Delta} y_{t\Delta}' y_{t\Delta} \hat{\beta}_{1\Delta}' - \hat{\gamma}_{1*} z_{t*}' z_{t*} \hat{\gamma}_{1*}')$$

und durch Einsetzen von (11.4) mit (11.3)

$$(11.7)\quad \sum_{t=1}^{T} \hat{u}_{1t}' \hat{u}_{1t} = \sum_{t=1}^{T} \hat{\beta}_{1\Delta} y_{t\Delta}' y_{t\Delta} \hat{\beta}_{1\Delta}' - \hat{\beta}_{1\Delta} y_{t\Delta}' z_{t*} (z_{t*}' z_{t*})^{-1} z_{t*}' y_{t\Delta} \hat{\beta}_{1\Delta}'$$

$$= \hat{\beta}_{1\Delta} \sum_{t=1}^{T} [y_{t\Delta}' y_{t\Delta} - y_{t\Delta}' z_{t*} (z_{t*}' z_{t*})^{-1} z_{t*}' y_{t\Delta}] \hat{\beta}_{1\Delta}'$$

$$= \hat{\beta}_{1\Delta} \cdot T \cdot W_{\Delta}(*) \ \hat{\beta}_{1\Delta}' \quad ,$$

wobei

$$(11.8)\quad W_{\Delta}(*) = \frac{1}{T} \sum_{t=1}^{T} [y_{t\Delta}' y_{t\Delta} - y_{t\Delta}' z_{t*} (z_{t*}' z_{t*})^{-1} z_{t*}' y_{t\Delta}].$$

$$= (M_{y_{t\Delta} y_{t\Delta}} - M_{y_{t\Delta} z_{t*}} M_{z_{t*} z_{t*}}^{-1} M_{z_{t*} y_{t\Delta}})$$

---

1) Diese Vereinfachung ergibt sich daraus, daß die Multiplikation von (11.4) mit $(z_{t*}' z_{t*})$ von rechts ergibt $\tilde{y}_t' z_{t*} = \hat{\gamma}_{1*} z_{t*}' z_{t*}$. Im übrigen sind alle Ausdrücke in (11.5) Skalare, wenn man nicht bis auf die Beobachtungswerte disaggregiert. Disaggregiert man jedoch, so ist darauf zu achten, daß die Summe der Quadrate der Residuen, dann als Produkt des Zeilenvektors mit dem Spaltenvektor der Residuen in dieser Reihenfolge ein Skalarprodukt ergeben muß, nämlich $\hat{u}_1 \hat{u}_1'$, wenn $\hat{u}_1$ als Zeilenvektor definiert ist.

Schätzen wir nun $\tilde{y}_t'$ nicht nur in Abhängigkeit von den in (11.2) enthaltenen vorherbestimmten Variablen sondern zugleich in Abhängigkeit von den aus (11.2) ausgeschlossenen vorherbestimmten Variablen z_{t**}' [1], so erhalten wir in gewohnter Weise[2] die Summe der Quadrate der Residuen dieser Regression aufgrund aller vorherbestimmten Variablen des Systems durch

$$(11.9)\quad \sum_{t=1}^{T} \hat{u}_{1t(K)}' \hat{u}_{1t(K)} = \sum_{t=1}^{T} (\tilde{y}_t' - \hat{\Upsilon}_1 z_t')(\tilde{y}_t - z_t \hat{\Upsilon}_1')$$

$$= \sum_{t=1}^{T} (\tilde{y}_t'\tilde{y}_t - \hat{\Upsilon}_1 z_t'\tilde{y}_t - \tilde{y}_t' z_t \hat{\Upsilon}_1' + \hat{\Upsilon}_1 z_t' z_t \hat{\Upsilon}_1')$$

$$= \sum_{t=1}^{T} (\tilde{y}_t' \tilde{y}_t - \hat{\Upsilon}_1 z_t' z_t \hat{\Upsilon}_1')$$

$$= \sum_{t=1}^{T} (\hat{\beta}_{1\Delta} y_{t\Delta}' y_{t\Delta} \hat{\beta}_{1\Delta}' - \hat{\beta}_{1\Delta} y_{t\Delta}' z_t (z_t' z_t)^{-1} \cdot z_t' y_{t\Delta} \hat{\beta}_{1\Delta}')$$

$$= \hat{\beta}_{1\Delta} \cdot T \cdot W_{\Delta(K)} \hat{\beta}_{1\Delta}',$$

wobei

$$(11.10)\quad W_{\Delta(K)} = \frac{1}{T} \sum_{t=1}^{T} [y_{t\Delta}' y_{t\Delta} - y_{t\Delta}' z_t (z_t' z_t)^{-1} z_t' y_{t\Delta}].$$

Bilden wir nun das Verhältnis der Streuung der Residuen der Regression aufgrund der in (11.2) enthaltenen vorherbestimmten Variablen zur Streuung der Residuen der Regression aufgrund aller vorherbestimmten Variablen des Systems, so

---

1) Insgesamt also in Abhängigkeit von $z_t' = \begin{bmatrix} z_{t*}' \\ z_{t**}' \end{bmatrix}$.

2) Entsprechend dem Vorgehen von (11.5) bis (11.8).

kürzt sich bei der Division von (11.6) bzw. (11.9) mit T, der Zahl der Beobachtungsperioden, T weg. Wir können also das Verhältnis der Summe der quadrierten Residuen dem Verhältnis der Streuung der Residuen gleichsetzen und erhalten

$$(11.11) \qquad \lambda = \frac{\sum_{t=1}^{T} \hat{u}_{1t}' \hat{u}_{1t}}{\sum_{t=1}^{T} \hat{u}_{1t(K)}' \hat{u}_{1t(K)}} = \frac{\hat{\beta}_{1\Delta} W_{\Delta(*)} \hat{\beta}_{1\Lambda}'}{\hat{\beta}_{1\Delta} W_{\Delta(K)} \hat{\beta}_{1\Lambda}'}$$

Da alle Ausdrücke in (11.8) und (11.10) quadratische Formen sind, können wir feststellen, daß der Zähler von (11.11) mindestens so groß ist wie der Nenner von (11.11), weil wir (11.10) erst erhalten, wenn wir von (11.8) positiv semidefinite[1] Ausdrücke subtrahieren. Daraus erhellt, daß $\lambda \geq 1$.

Das Prinzip des minimalen Streuungsverhältnisses beinhaltet somit, diejenige Schätzung $\hat{\beta}_{1\Lambda}$ zu suchen, die die Streuung im Zählerausdruck in (11.11) nach Möglichkeit auf die Streuung im Nennerausdruck reduziert: Es wird also der kleinste Skalar $\hat{\lambda}$ gesucht unter Berücksichtigung der Beobachtungswerte insbesondere der aus (11.2) ausgeschlossenen vorherbestimmten Variablen. Koopmans und Hood[2] beleuchten denselben Aspekt von einer anderen Seite: In der Strukturgleichung (11.2) kommen die vorherbestimmten Variablen z_{t**} nicht vor. Daher soll durch die Berücksichtigung dieser aus (11.2) ausgeschlossenen vorherbestimmten Variablen z_{t**}, deren Beobachtungswerte aber in unserer Stichprobe vom Umfang T enthalten sind, nur eine minimale Verbesserung der Erklärung der Variablen $\tilde{y}_t'$ erfolgen im Vergleich zur Außerachtlassung von z_{t**} bei der Erklärung dieser Variablen $\tilde{y}_t'$.

---

1) Vgl. Menges (1961): a.a.O., S. 224.

2) Vgl. zu diesem Absatz: Koopmans u. Hood: a.a.O., S. 169.

Gleichung (11.11) ist identisch mit Gleichung (10.61), wie man durch Vergleich von $W_{\Delta(*)}$ (11.8) mit $W^*_{\Delta,\Delta}$ (10.58) und von $W_{\Delta(K)}$ (11.10) mit $W_{\Delta,\Delta}$ (10.54) feststellen kann. Die Methode der größten Dichte bei beschränkter Information nach Koopmans und Hood führt somit zu den gleichen Schätzergebnissen wie das Prinzip des minimalen Streuungsverhältnisses.

Wenn wir nun (11.11) umformen zu

$$\hat{\beta}_{1\Delta} W_{\Delta(*)} \hat{\beta}'_{1\Delta} = \lambda \hat{\beta}_{1\Delta} W_{\Delta(K)} \hat{\beta}'_{1\Delta} \tag{11.12}$$

und bezüglich $\hat{\beta}_{1\Delta}$ differenzieren, so erhalten wir im Minimum

$$\frac{\partial \hat{\beta}_{1\Delta} W_{\Delta(*)} \hat{\beta}'_{1\Delta}}{\partial \hat{\beta}_{1\Delta}} = \frac{\partial \lambda \hat{\beta}_{1\Delta} W_{\Delta(K)} \hat{\beta}'_{1\Delta}}{\partial \hat{\beta}_{1\Delta}} \tag{11.13}$$

$$= \hat{\beta}_{1\Delta} W_{\Delta(*)} = \lambda \hat{\beta}_{1\Delta} W_{\Delta(K)} = 0 \quad ^{1)}$$

und somit gilt

$$\hat{\beta}_{1\Delta} (W_{\Delta(*)} - \lambda W_{\Delta(K)}) = 0 \quad ^{1)} \tag{11.14}$$

Damit wir eine nicht triviale Schätzung $\hat{\beta}_{1\Delta}$ erhalten, muß der Klammerausdruck singulär sein, d.h. es muß gelten

$$|W_{\Delta(*)} - \lambda W_{\Delta(K)}| = 0 \quad ^{2)} \tag{11.15}$$

Wir suchen dann die kleinste[3)] Wurzel $\hat{\lambda}$ dieser charakteristischen Gleichung, setzen sie wieder in (11.15) ein und berechnen dann $\hat{\beta}_{1\Delta}$.

---

1) 0 steht hier für den Zeilenvektor mit G^Δ Nullen.

2) 0 steht hier für den Skalar Null.

3) Vgl. S. 112.

$\hat{\gamma}_{1*}$ können wir schließlich unter Zuhilfenahmen von (11.4) und (11.3) berechnen mit

$$\hat{\gamma}_{1*} = \hat{\beta}_{1\Delta}\, y_{t\Delta}'\, z_{t*}(z_{t*}'\, z_{t*})^{-1}. \tag{11.16}$$

Da das Prinzip des minimalen Streuungsverhältnisses zu den gleichen Schätzungen führt wie die Methode der größten Dichte bei beschränkter Information nach Koopmans und Hood, können wir im zweiten Teil dieser Arbeit auf eine kritische Erörterung der Schätzungen nach dem Prinzip des minimalen Streuungsverhältnisses verzichten.

12. Die Schätzungen der k-Klasse[1].

Wir gehen aus von einer Gleichung

$$y_1 = y_2 \beta_2' + X_* \gamma_*' + u \tag{12.1}$$

wobei

$$y_1 = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1T} \end{bmatrix}, \quad y_2 = \begin{bmatrix} y_{21} & \cdots & y_{G\Delta_1} \\ \vdots & & \vdots \\ y_{2T} & \cdots & y_{G\Delta_T} \end{bmatrix}, \quad \beta_2' = \begin{bmatrix} \beta_2 \\ \vdots \\ \beta_{G\Delta} \end{bmatrix}, \tag{12.2}$$

$$X_* = \begin{bmatrix} x_{11} & \cdots & x_{K*1} \\ \vdots & & \vdots \\ x_{1T} & \cdots & x_{K*T} \end{bmatrix},$$

$$\gamma_*' = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_{K*} \end{bmatrix} \quad \text{und} \quad u = \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix}.$$

---

1) Vgl. hierzu: Nagar, A.L.: The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. Econometrica, 27(1959), S. 575-595.

Theil, H. (1961): a.a.O., S. 225-232.

Johnston, J.: a.a.O., S. 260-263.

Durch Einsetzen der reduzierte-Form-Schätzung von y_2[1], nämlich von $\hat{y}_2 = y_2 - V$, gelangen wir zu

$$(12.3) \quad y_1 = (y_2 - V)\beta_2' + X_*\gamma_*' + (u + V\beta_2')$$

Wir sehen, daß (12.1) und (12.3) identisch sind, da $-V\beta_2'$ und $+V\beta_2'$ sich gegenseitig aufheben.

Für (12.3) können wir auch schreiben

$$(12.4) \quad y_1 = z \begin{bmatrix} \beta_2' \\ \gamma_*' \end{bmatrix} + w$$

wobei

$$(12.5) \quad z = [(y_2 - V) \quad X_*] \quad \text{und} \quad w = (u + V\beta_2')$$

Unter den Voraussetzungen, daß $K^{**} \geq G^{\Delta} - 1$ und daß $E(w) = 0$, wobei 0 der Spaltenvektor mit T Nullen ist, und $E(z'w) = 0$, wobei 0 der Spaltenvektor mit $(G^{\Delta} - 1 + K^*)$ Nullen ist, schätzen wir nun (12.4) mit der gewöhnlichen Methode der kleinsten Quadrate und erhalten

---

1) Es ist nämlich die reduzierte-Form-Schätzung von y_2

$$y_2 - V = \hat{y}_2 = X\hat{\Pi} = X(X'X)^{-1}X'y_2$$

wobei $X = [X_* \quad X_{**}]$ auch die Matrix der aus Gleichung (12.1) ausgeschlossenen vorherbestimmten Variablen X_{**} enthält. X ist eine $(T \times K)$ Matrix und $\hat{\Pi}$ ist die zugehörige Matrix der reduzierte-Form-Parameterschätzungen und von der Ordnung $[K \times (G^{\Delta} - 1)]$. V ist die Matrix der reduzierte-Form-Residuen und von der Ordnung $[T \times (G^{\Delta}-1)]$. Wir erhalten für V

$$V = y_2 - \hat{y}_2 = y_2 - X(X'X)^{-1}X'y_2 = [I_T - X(X'X)^{-1}X']y_2$$

$$(12.6)\quad \begin{bmatrix} \hat{\beta}_2' \\ \hat{\gamma}_*' \end{bmatrix} = (z'z)^{-1} z'y_1$$

$$= \begin{bmatrix} (y_2-V)'(y_2-V) & (y_2-V)'X_* \\ X_*'(y_2-V) & X_*' X_* \end{bmatrix}^{-1} \begin{bmatrix} (y_2-V)' \\ x_*' \end{bmatrix} y_1$$

Da $X_* = X\begin{bmatrix} I \\ 0 \end{bmatrix}$, wobei I die Einheitsmatrix der Ordnung K* und 0 die Nullmatrix der Ordnung (K** x K*) sind, ist

$$(12.7)\quad X_*' V = \begin{bmatrix} I \\ 0 \end{bmatrix}' X'[I_T - X(X'X)^{-1}X']y_2 = \begin{bmatrix} I \\ 0 \end{bmatrix}' (X'-X')y_2 = 0$$

und entsprechend ist $V'X_* = 0$.

Weiter gilt

$$(12.8)\quad (y_2 - V)'(y_2 - V) = y_2' y_2 - y_2'V - V'y_2 + V'V$$

und wir erhalten für

$$(12.9)\quad \begin{aligned} V'y_2 &= V'[V + X(X'X)^{-1}X'y_2] \\ &= V'V + y_2'[I_T - X(X'X)^{-1}X'][X(X'X)^{-1}X'y_2] \\ &= V'V + y_2' X(X'X)^{-1}X'y_2 - y_2'X(X'X)^{-1}X'y_2 \\ &= V'V \end{aligned}$$

Entsprechend erhalten wir für $y_2' V = V'V$, so daß folgt

$$(12.10)\quad (y_2 - V)'(y_2 - V) = y_2' y_2 - V'V \ .$$

Somit erhalten wir für (12.6)

$$(12.11)\qquad \begin{bmatrix} \hat{\beta}_2' \\ \hat{\gamma}_*' \end{bmatrix} = \begin{bmatrix} y_2' y_2 - V'V & y_2' X_* \\ X_*' y_2 & X'_* X_* \end{bmatrix}^{-1} \begin{bmatrix} (y_2 - V)' \\ X_*' \end{bmatrix} y_1$$

Die Schätzungen der k-Klasse erhalten wir nun, wenn wir V'V bzw. V' mit k multiplizieren:

$$(12.12)\qquad \begin{bmatrix} \hat{\beta}_2' \\ \hat{\gamma}_*' \end{bmatrix}_k = \begin{bmatrix} y_2' y_2 - kV'V & y_2' X_* \\ X_*' y_2 & X_*' X_* \end{bmatrix}^{-1} \begin{bmatrix} (y_2 - kV)' \\ X_*' \end{bmatrix} y_1$$

Für k = 0 erhalten wir dann

$$(12.13)\qquad \begin{bmatrix} \hat{\beta}_2' \\ \hat{\gamma}_*' \end{bmatrix}_{k=0} = \begin{bmatrix} y_2' y_2 & y_2' X_* \\ X_*' y_2 & X_*' X_* \end{bmatrix}^{-1} \begin{bmatrix} y_2' \\ X_*' \end{bmatrix} y_1$$

d.h. die Schätzung nach der gewöhnlichen Methode der kleinsten Quadrate, direkt auf Gleichung (12.1) angewandt.

Für k = 1 erhalten wir

$$(12.14)\qquad \begin{bmatrix} \hat{\beta}_2' \\ \hat{\gamma}_*' \end{bmatrix}_{k=1} = \begin{bmatrix} y_2' y_2 - V'V & y_2' X_* \\ X_*' y_2 & X_*' X_* \end{bmatrix}^{-1} \begin{bmatrix} (y_2 - V)' \\ X_*' \end{bmatrix} y_1$$

d.h. die Schätzung nach der zweistufigen Methode der kleinsten Quadrate[1].

1) Vgl. (5.34) mit (12.14) bzw. (12.11).

Nun untersuchen wir noch die Identität[1] mit der Schätzung nach der Methode der größten Dichte bei beschränkter Information nach Koopmans und Hood (die ihrerseits identisch ist mit dem Prinzip des minimalen Streuungsverhältnisses).

Eine Umformung von (12.12) ergibt

$$(12.15)\quad \begin{bmatrix} y_2' y_2 - kV'V & y_2' X_* \\ X_*' y_2 & X_*' X_* \end{bmatrix} \begin{bmatrix} \hat{\beta}_2' \\ \hat{\gamma}_*' \end{bmatrix}_k = \begin{bmatrix} (y_2 - kV)' \\ X_*' \end{bmatrix} y_1$$

Die obere Gleichung in (12.15) lautet ausmultipliziert:

$$(2.16)\quad (y_2' y_2 - kV'V)(\hat{\beta}_2')_k + y_2' X_* (\hat{\gamma}_*')_k = (y_2 - kV)' y_1$$

Wir verwenden nun

$$(12.17)\quad V = y_2 - X(X'X)^{-1}X'y_2$$
$$= [I_T - X(X'X)^{-1}X']y_2 = A\, y_2 ,$$

wobei

$$(12.18)\quad A = [I_T - X(X'X)^{-1}X']$$

eine symmetrische, idempotente[2] Matrix ist, d.h. $A = A'$ und $A\,A = A$.

Wir erhalten dann für

$$(12.19)\quad V'V = y_2' A'A y_2 = y_2' A\, y_2$$

---

1) Vgl. hierzu besonders Johnston: a.a.O., S. 260-262.
2) Vgl. Zurmühl: a.a.O., S. 114.

und für

$$V'y_1 = y_2' A'y_1 = y_2' A y_1. \tag{12.20}$$

(12.19) und (12.20), in (12.16) eingesetzt, ergeben

$$(y_2' y_2 - k\, y_2' A y_2)(\hat{\beta}_2')_k + y_2' X_*(\hat{\gamma}_*')_k - y_2' y_1 + k\, y_2' A y_1 = 0 \tag{12.21}$$

Die untere Gleichung in (12.15), nach $(\hat{\gamma}_*')_k$ aufgelöst, lautet

$$(\hat{\gamma}_*')_k = (X_*' X_*)^{-1}[X_*' y_1 - X_*' y_2(\hat{\beta}_2')_k]. \tag{12.22}$$

Der zweite Ausdruck auf der linken Seite von (12.21) lautet dann

$$y_2' X_*(\hat{\gamma}_*')_k = y_2' X_*(X_*' X_*)^{-1}[X_*' y_1 - X_*' y_2(\hat{\beta}'_2)_k] \tag{12.23}$$

Verwenden wir (12.23) in (12.21), so erhalten wir

$$(y_2' y_2 - k\, y_2' A y_2)(\hat{\beta}_2')_k + y_2' X_*(X_*' X_*)^{-1}X_*' y_1 - y_2' X_*(X_*' X_*)^{-1}X_*' y_2(\hat{\beta}_2')_k - y_2' y_1 + k\, y_2' A y_1 = 0 \tag{12.24}$$

Durch Umgruppierung gelangen wir zu

$$[y_2' y_2 - y_2' X_*(X_*' X_*)^{-1}X_*' y_2](\hat{\beta}_2')_k - k\, y_2' A y_2(\hat{\beta}_2')_k - [y_2' y_1 - y_2' X_*(X_*' X_*)^{-1}X_*' y_1] + k\, y_2' A y_1 = 0 \tag{12.25}$$

Definieren wir nun

$$(12.26) \qquad [I_T - X_*(X_*' X_*)^{-1}X_*'] = A_*$$

so erhalten wir aus (12.25) nach einer weiteren Umformung

$$(12.27) \quad (y_2' A_* y_2 - k\, y_2' A y_2)(\hat{\beta}_2')_k - (y_2' A_* y_1 - k y_2' A y_1) = 0$$

oder

$$(12.28) \quad (y_2' A_* - k\, y_2' A)\,[y_1 \quad y_2] \begin{bmatrix} -1 \\ (\hat{\beta}_2')_k \end{bmatrix} = 0$$

Nun ist aber

$$(12.29) \qquad [y_1 \quad y_2] = Y_\Delta$$

die Matrix aller in (12.1) auftretenden gemeinsam abhängigen Variablen und

$$(12.30) \qquad \begin{bmatrix} -1 \\ (\hat{\beta}_2')_k \end{bmatrix} = (\hat{\beta}_{1\Delta}')_k$$

der zugehörige Koeffizientenvektor.

Wegen (12.29) und (12.30) können wir also schreiben

$$(12.31) \quad (y_2' A_* Y_\Delta - k\, y_2' A Y_\Delta)(\hat{\beta}_{1\Delta}')_k = 0$$

Wir behaupten nun, Gleichung (12.31) sei identisch mit der Gleichung (11.14) des Prinzips des minimalen Streuungsverhältnisses:

$$(11.14) \qquad \hat{\beta}_{1\Delta}(W_{\Delta(*)} - \lambda\, W_{\Delta(K)}) = 0$$

oder, da $W_{\Delta(*)}$ (11.8) und $W_{\Delta(K)}$ (11.11) symmetrisch sind,

$$(11.15) \qquad (W_{\Delta(*)} - \lambda W_{\Delta(K)}) \hat{\beta}'_{1\Delta} = 0 ,$$

wobei

$$(11.8) \quad W_{\Delta(*)} = (M_{y_{t\Delta}y_{t\Delta}} - M_{y_{t\Delta}z_{t*}} M^{-1}_{z_{t*}z_{t*}} M_{z_{t*}y_{t\Delta}})$$

$$= \frac{1}{T} \sum_{t=1}^{T} [y_{t\Delta}'y_{t\Delta} - y'_{t\Delta}z_{t*}(z'_{t*}z_{t*})^{-1}z'_{t*}y_{t\Delta}]$$

$$= [y'_{\Delta} y_{\Delta} - y'_{\Delta} z_{*}(z'_{*} z_{*})^{-1}z'_{*} y_{\Delta}]$$

$$= Y'_{\Delta} A_{*} Y_{\Delta} ,$$

da X_* in (12.26) nichts anderes ist als z_*, und wobei

$$(11.10) \quad W_{\Delta(K)} = \frac{1}{T} \sum_{t=1}^{T} [y_{t\Delta}'y_{t\Delta} - y_{t\Delta}' z_t(z_t' z_t)^{-1}z_t' y_{t\Delta}]$$

$$= [y'_{\Delta} y_{\Delta} - y'_{\Delta} z(z'z)^{-1}z'y_{\Delta}] = Y'_{\Delta} A Y_{\Delta} ,$$

da X in (12.18) nichts anderes ist als z.

Wir erhalten aus (11.16) somit

$$(12.32) \qquad (Y'_{\Delta} A_{*} Y_{\Delta} - \lambda Y'_{\Delta} A Y_{\Delta})\hat{\beta}'_{1\Delta} = 0$$

Verwenden wir (12.29) , so erhalten wir aus (12.32)

$$(12.33) \quad (y'_1 A_{*} Y_{\Delta} - \lambda y'_1 A Y_{\Delta})(\hat{\beta}'_{1\Delta}) = 0 \ ^{1)}$$

---

1) Diese Gleichung ist nur von formalem Interesse, weil wir durch die a-priori-Normierung, d.h. Auflösung der Gleichung (12.1) nach y_1, den Parameter β_1 bereits Eins gleichgesetzt haben und daher diesen nicht mehr schätzen müssen.

und

$$(12.34) \qquad (y_2' A_* Y_\Delta - \lambda\, y_2' A Y_\Delta)(\hat{\beta}_{1\Delta}') = 0$$

Gleichung (12.34) ist identisch mit Gleichung (12.31), wenn $k = \lambda$, d.h. für $k = \lambda$ ist die k-Klassen-Schätzung die Schätzung nach dem Prinzip des minimalen Streuungsverhältnisses.[1)]

Um noch die Identität von $(\hat{\gamma}_*')k = \lambda$ und der Parameterschätzung $\hat{\gamma}_{1*}'$ nach dem Prinzip des minimalen Streuungsverhältnisses zu beweisen, betrachten wir letztere, welche aus (11.16) hervorgeht:

$$(12.35) \qquad \hat{\gamma}_{1*}' = (z_{t*}' z_{t*})^{-1} z_{t*}' y_{t_\Delta} \hat{\beta}_{1\Delta}'$$

Dies ist gleich dem Negativen[2)] von (12.22), da wir aus (12.29) und (12.30)

$$(12.36) \qquad Y_\Delta(\hat{\beta}_{1\Delta}')_k = [-y_1 + y_2(\hat{\beta}_2')_k]$$

herleiten können und y_{t_Δ} bei der Schätzung zu Y_Δ aggregiert wird.

Wir stellen also fest, daß die Schätzungen der k-Klasse Schätzungen nur für eine einzige Gleichung eines Gleichungssystems sind und daher im allgemeinen nicht zur simultanen Schätzung aller Gleichungen eines Gleichungssystems geeignet sind.

---

1) Dann ist die k-Klassenschätzung für $k = \lambda$ auch die Schätzung nach der Methode der größten Dichte bei beschränkter Information nach Koopmans und Hood.

2) Im 11. Kapitel hatte γ_{1*} ein negatives Vorzeichen [Vgl. (11.2)], im 12. Kapitel hat γ_*' ein positives Vorzeichen [Vgl. (12.3)].

Da für $k = 0$, $k = 1$ bzw. $k = \lambda$ die Schätzungen der k-Klasse als gewöhnliche Kleinst-Quadrat-Schätzungen, Schätzungen der zweistufigen Methode der kleinsten Quadrate bzw. Schätzungen der Methode der größten Dichte bei beschränkter Information nach Koopmans und Hood bekannt sind, wird im folgenden auf eine weitere summarische Diskussion der k-Klassen-Schätzungen verzichtet, da die Kritik sich auf die erwähnten drei k-Klassen Schätzungen konzentrieren wird.

Im übrigen sei noch folgendes - meiner Ansicht nach interessantes - Detail erwähnt:

Geht man anstelle von Gleichung (12.3) von folgender Gleichung aus

$$(12.37) \qquad y_1 = (y_2 - kV)\beta_2' + X_* \gamma_*' + (u + k V \beta_2'),$$

so gelangt man wegen[1)]

$$(12.38) \qquad k V' y_2 = k y_2' V = k V'V k$$

anstatt zu (12.12) zu der Schätzung

$$(12.39) \qquad \begin{bmatrix} \hat{\beta}_2' \\ \hat{\gamma}_*' \end{bmatrix}_{kk} = \begin{bmatrix} y_2'y_2 - kV'Vk & y_2' X_* \\ X_*' y_2 & X_*' X_* \end{bmatrix}^{-1} \begin{bmatrix} (y_2 - k V)' \\ X_*' \end{bmatrix} y_1$$

welche m.E. ebenfalls zur Doppel-k-Klasse Nagars gehören muß[2)].

1) Vgl. auch (12.6) und (12.9).

2) Vgl. hierzu die Bemerkung von Theil (1961): a.a.O., S. 354.

13. Three-Pass Least Squares

Taylor und Wilson[1] präsentierten 1964 diese Methode, die speziell für Modelle mit verzögerten endogenen Variablen geeignet ist.

Das betrachtete Modell besteht aus einer Gleichung

$$(13.1) \qquad y_t = \alpha\, y_{t-1} + \beta\, x_t + u_t \qquad (t=2,\ldots,T)$$

mit einer unverzögerten endogenen Variablen y_t, einer verzögerten endogenen Variablen y_{t-1}, einer exogenen Variablen x_t[2], der latenten Variablen u_t und den Parametern α und β.

Die latente Variable der Periode t sei in folgender Weise mit der latenten Variablen der Vorperiode verknüpft

$$(13.2) \qquad u_t = \lambda\, u_{t-1} + \varepsilon_t,$$

wobei λ eine unbekannte Konstante ist.

Hierbei wird angenommen, daß ε_t eine reelle stochastische Variable ist, die unabhängig verteilt ist mit endlicher, zeitlich konstanter Streuung $\sigma = \sigma_t$ für $t=2,\ldots,T$. Außerdem wird angenommen, daß x_t eine nicht stochastische Variable ist.

---

1) Taylor, L.D. u. Wilson, T.A.: Three-Pass Least Squares: A method for estimating models with a lagged dependent variable. The Review of Economics and Statistics, 46 (1964), S. 329-346, s. bes. S. 330-331.

2) Die Werte y_t, y_{t-1} und x_t seien die Abweichung der ursprünglichen Werte dieser Variablen von ihrem Mittelwert, so daß in Gleichung (13.1) kein absolutes Glied mehr auftritt. $(t=2,\ldots,T)$, obwohl von Taylor und Wilson nicht besonders erwähnt, gilt auch für die folgenden Größen, die den Index t tragen.

Wenn wir (13.2) in (13.1) einsetzen, erhalten wir

$$(13.3) \qquad y_t = \alpha y_{t-1} + \beta x_t + \lambda u_{t-1} + \epsilon_t$$

Das Schätzverfahren von Taylor und Wilson besteht, wie der Name besagt, aus drei Schritten.

Während des ersten Schrittes wird die Gleichung (13.1) mit der gewöhnlichen Methode der kleinsten Quadrate geschätzt. Als Ergebnis dieser Schätzung erhält man

$$(13.4) \qquad y_t = a_1 y_{t-1} + b_1 x_t + v_t$$

Hierbei sind a_1 die Schätzung des Parameters α durch den ersten Schritt und b_1 die Schätzung des Parameters β durch den ersten Schritt und v_t das Residuum für die Periode t, das man durch Einsetzen der Beobachtungswerte von y_t, y_{t-1} und x_t in die Schätzfunktion (13.4) des ersten Schrittes berechnen kann.

Wir nehmen nun an, daß Wahrscheinlichkeitsgrenzen für die Parameterschätzungen a_1 und b_1 und für das Residuum v_t bestehen. Dann können wir den Wahrscheinlichkeitslimes für v_t unter Verwendung von (13.4) bezeichnen mit

$$(13.5) \qquad \operatorname{plim} v_t = y_t - (\operatorname{plim} a_1)\, y_{t-1} - (\operatorname{plim} b_1) x_t$$

Wenn wir die Differenz zwischen dem Wahrscheinlichkeitslimes des Residuums v_t und der "wahren" aber unbekannten, konstanten latenten Variablen u_t bilden, so erhalten wir unter Berücksichtigung von (13.1)

$$\begin{aligned} (13.6) \qquad (\operatorname{plim} v_t) - u_t &= y_t - (\operatorname{plim} a_1) y_{t-1} - (\operatorname{plim} b_1) x_t \\ &\quad - y_t + \alpha y_{t-1} + \beta x_t \\ &= (\alpha - \operatorname{plim} a_1) y_{t-1} + (\beta - \operatorname{plim} b_1) x_t . \end{aligned}$$

Dann ist

(13.7) $\text{plim } v_t = u_t + (\alpha - \text{plim } a_1)y_{t-1} + (\beta - \text{plim } b_1)x_t.$

Der bias, der zwischen den "wahren" Parametern α bzw. β und dem Wahrscheinlichkeitslimes ihrer Schätzungen durch den ersten Schritt a_1 bzw. b_1 bestehe, sei nun eine linear-homogene Funktion der Konstanten λ, die in (13.2) auftritt[1]. Dann bestehen die Beziehungen

(13.8) $\text{plim } a_1 = \alpha + \lambda\ \delta$

und

(13.9) $\text{plim } b_1 = \beta + \lambda\ \varsigma$

wobei δ und ς unbekannte Konstanten sind.

Wenn x_t nicht autokorreliert ist, wird $\varsigma = 0$ sein, weil $\text{plim } b_1 = \beta$. Dann wird der Ausdruck $(\beta - \text{plim } b_1)x_t$ in (13.7) fortfallen. Durch Einsetzen von (13.8) in (13.7) gelangen wir in diesem Fall zu

(13.10) $\text{plim } v_t = u_t - \lambda\ \delta\ y_{t-1}$

v_t ist also unter den gemachten Annahmen eine asymptotisch verzerrte, also nicht konsistente, Schätzung von u_t, wobei die Verzerrung von y_{t-1} abhängt. Wenn λ und δ bekannt wären, könnten wir $v_{t-1} + \lambda\ \delta\ y_{t-2}$ als konsistente Schätzung für u_{t-1} in (13.3) einsetzen und könnten dann mit (13.3) bereits beim zweiten Schritt konsistente Schätzungen von α und β erzielen, wenn wir (13.3) danach[2] mit der gewöhnlichen Methode der kleinsten Quadrate schätzen wür-

---

1) Warum die asymptotische Verzerrung gerade proportional zum Autoregressionskoeffizienten λ ist, wird von Taylor und Wilson nicht erklärt. Es kann sich daher nur um eine erste Approxiamtion handeln.

2) Nach diesem Einsetzen hätte (13.3) die Gestalt von (13.11).

den. Da wir jedoch λ und δ nicht kennen, können wir lediglich $v_{t-1} + \lambda \delta y_{t-2}$ als konsistente, aber uns noch unbekannte Schätzung für u_{t-1} in (13.3) einsetzen und erhalten

$$(13.11) \quad y_t = \alpha y_{t-1} + \beta x_t + \lambda v_{t-1} + \lambda^2 \delta y_{t-2} + \epsilon_t$$

Da wir angenommen haben, daß x_t nicht autokorreliert ist[1], können wir die bereits ermittelte konsistente Schätzung b_1 für β in (13.11) einsetzen und erhalten

$$(13.12) \quad y_t - b_1 x_t = \alpha y_{t-1} + \lambda v_{t-1} + (\lambda^2\delta) y_{t-2} + \epsilon_t$$

Während des zweiten Schrittes wird (13.12) mit der gewöhnlichen Methode der kleinsten Quadrate geschätzt. y_t und $b_1 x_t$ wurden zusammengefaßt, weil sie wegen der Kenntnis der im ersten Schritt erhaltenen Parameterschätzung b_1 die einzigen Ausdrücke in (13.12) sind, deren Wert man schon kennt.

Als Ergebnis dieser Schätzung erhält man

$$(13.13) \quad y_t - b_1 x_t = a_2 y_{t-1} + l_2 v_{t-1} + (l_2^2 d_2) y_{t-2} + w_t,$$

wobei der Index 2 die Parameterschätzungen des 2. Schrittes kennzeichnet und w_t das Residuum der Schätzung des 2. Schrittes ist.

Wäre β im zweiten Schritt neu geschätzt worden, so hätten wir ein b_2 erhalten, dessen Streuung für endliche Stichproben größer wäre als die Streuung von b_1. Dies resultiert

---

1) Siehe (13.10) im Unterschied zu (13.7).

aus der um drei geringeren Zahl der Freiheitsgrade von (13.13) im Vergleich zu (13.3), denn die Aufnahme von y_{t-2} und v_{t-1} in (13.13) zwingt zum Verzicht auf die erste Beobachtung für t = 2, da nur ein t-1, nicht jedoch ein t-2 existiert. Außerdem sind die Koeffizienten dieser beiden neuen Variablen zu schätzen.

Die konsistenten Schätzungen b_1 von ß und a_2 von α werden nun in (13.1) eingesetzt, um zu einer konsistenten Schätzung von u_t zu gelangen. Wir erhalten also

$$(13.14) \qquad \hat{u}_t = y_t - a_2 y_{t-1} - b_1 x_t$$

Wenn wir die Schätzung der latenten Variablen um eine Periode verzögern und in (13.3) einsetzen, ergibt sich

$$(13.15) \qquad y_t = \alpha y_{t-1} + ß x_t + \lambda \hat{u}_{t-1} + \epsilon^*_t,$$

wobei $\epsilon^*_t = \epsilon_t + \lambda e_{t-1}$, wenn $e_{t-1} = u_{t-1} - \hat{u}_{t-1}$.[1)]

Während des dritten Schrittes wird die Gleichung (13.15) mit der gewöhnlichen Methode der kleinsten Quadrate geschätzt.

Da $\hat{u}_{t-1}$ asymptotisch gegen u_{t-1} konvergiert, wird auch ϵ^*_t asymptotisch gegen ϵ_t konvergieren.

Als Ergebnis dieser dritten Schätzung erhält man

$$(13.16) \qquad y_t = a_3 y_{t-1} + b_3 x_t + l_3 \hat{u}_{t-1} + \eta_t$$

---

1) Diese Beziehung zwischen ϵ^*_t und ϵ_t wird von Taylor und Wilson nicht besonders herausgearbeitet, obwohl sie für die asymptotische Betrachtung von ϵ^*_t von Bedeutung ist.

mit den konsistenten Parameterschätzungen a_3, b_3 und l_3. η_t ist das Residuum der Schätzung des 3. Schrittes.

Da dieses Schätzverfahren speziell für eine Gleichung mit nur einer verzögerten endogenen Variablen konzipiert ist, kann auf einen Vergleich mit anderen Schätzverfahren verzichtet werden, die nicht nur für den Spezialfall verzögerter Variablen konsistente Schätzungen liefern. Gleichwohl mag dieses Verfahren in einigen ökonometrischen Anwendungsfällen, in denen der Parameter einer verzögerten endogenen Variablen hochsignifikant ist, von Bedeutung sein.

14. Die Momentenmethode[1)]

Diese Methode wurde von Karl Pearson[2)] vorgeschlagen. "Die Methode besteht darin, eine geeignete Anzahl der Stichprobenelemente den entsprechenden Momenten der Verteilung, welche Funktionen der unbekannten Parameter sind, gleichzusetzen. Durch Berücksichtigung so vieler Momente, als Parameter zu schätzen sind, und Lösung der sich ergebenden Gleichungen nach den Parametern werden Schätzungen der letzteren erzielt."[3)]

Da die so erhaltenen Schätzungen aus den n Stichprobenwerten Funktionen der Stichprobenelemente sind, weisen sie die Eigenschaften der Stichprobenverteilung auf. Unter ziemlich allgemeinen Bedingungen wird daher die Verteilung einer solchen Schätzung asymptotisch normal für große n sein und der Mittelwert der Schätzung wird vom wahren Wert des Parameters mit einer Größe der Ordnung $\frac{1}{n}$ abweichen. Die Standardabweichung wird dann asymptotisch von der Gestalt $\frac{c}{\sqrt{n}}$ sein. Durch einen Korrekturfaktor kann man oft die Verzerrung einer solchen Schätzung beseitigen und er-

---

1) Cramér: a.a.O., S. 497-498.

2) Pearson, K.: Contributions to the mathematical theory of evolution. I. Philosophical Transactions of the Royal Society, Series A, 185(1894), S. 71 ff, ders.: Contributions to the mathematical theory of evolution. IV.: On the probable errors of frequency constants and on the influence of random selection on variation and correlation. Philosophical Transactions of the Royal Society, Series A, 191(1898), S. 229 ff, ders.: On the systematic fitting of curves to observations and measurements. Biometrika, 1(1902), S. 265-303 und Biometrika, 2(1902), S. 1-23.

3) Cramér: a.a.O., S. 497. Übersetzung des Verfassers.

hält dann eine erwartungstreue Schätzung. Unter allgemeinen Bedingungen wird diese Methode also asymptotisch effizient sein. R.A. Fisher[1] zeigte, daß im Falle mehrerer Parameter die Schätzungen der Momentenmethode nicht effizient sind.

Für die Normalverteilung lauten die Schätzungen der unbekannten Parameter μ und σ^2 mit der Momentenmethode $\hat{\mu}$ und $\hat{\sigma}^2$. Durch Beseitigung der Verzerrung erhalten wir die erwartungstreuen und asymptotisch effizienten Schätzungen $\hat{\mu}$ und $\frac{n}{n-1}\hat{\sigma}^2$. Für die anderen Verteilungen des Systems von Pearson (Pearsontypen) gilt diese asymptotische Effizienz nicht, wie bereits Fisher[1] gezeigt hat.

Wie Schneeweiß[2] gezeigt hat, führt die Momentenmethode zu identischen Schätzungen wie die Methode der kleinsten Quadrate, wenn für das 1. Moment, d.h. den Mittelwert, der Verteilung angenommen wird:

$$(14.1) \qquad E\, u_t = 0,$$

wobei u_t die latente Variable ist und die Kovarianz zwischen den latenten Variablen u_t und den exogenen Variablen x_t in der wahren Struktur gleich Null ist:

$$(14.2) \qquad E(u_t\, x_t) = 0.$$

Die entsprechenden Ausprägungen der Stichprobe lauten dann für die Residuen e_t

$$(14.3) \qquad \frac{1}{T} \sum_{t=1}^{T} e_t = 0$$

---

1) Fisher, R.A.: On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society, Series A, 222(1921), S. 309 ff.

2) Schneeweiß: a.a.O., S. 22.

und

$$(14.4) \qquad \frac{1}{T}\sum_{t=1}^{T} e_t x_t = 0.$$

Aus (14.3) erhalten wir wegen

$$(14.5) \qquad e_t = y_t - \hat{y}_t = y_t - \hat{\alpha} - \hat{\beta} x_t$$

die erste Normalgleichung[1] der Methode der kleinsten Quadrate

$$(14.6) \qquad \sum_{t=1}^{T} (y_t - \hat{\alpha} - \hat{\beta} x_t) = 0.$$

Entsprechend erhalten wir aus (14.4) und (14.5) die zweite Normalgleichung[2] der Methode der kleinsten Quadrate

$$(14.7) \qquad \sum_{t=1}^{T} (y_t - \hat{\alpha} - \hat{\beta} x_t) x_t = 0.$$

Damit führt also unter den Voraussetzungen (14.1) und (14.2) die Momentenmethode zu identischen Schätzungen wie die gewöhnliche Methode der kleinsten Quadrate, so daß im folgenden bei der kritischen Erörterung der Methode der kleinsten Quadrate implizit die Momentenmethode mitbehandelt wird.

---

1) Vgl. (2.5).
2) Vgl. (2.6).

15. Die χ^2-Minimum-Methode[1]

Wir gehen von folgender Wahrscheinlichkeitsfunktion

$$(15.1) \qquad P(X;\ \theta_1,\ \ldots,\ \theta_s)$$

aus, deren mathematische Form uns bekannt, deren Parameter $\theta_1, \ldots, \theta_s$ uns jedoch unbekannt sind. X stelle die Menge der Stichprobenelemente dar. Die Anzahl der Stichprobenelemente sei n, unsere Hypothese lautet nun, daß die Stichprobe aus einer Grundgesamtheit, welche dieser Wahrscheinlichkeitsfunktion gehorcht, gezogen wurde.

Wir nehmen an, daß unsere Stichprobe X in r voneinander unabhängige Gruppen $X_1, \ldots, X_r$ aufgeteilt werden kann, deren empirische Häufigkeiten wir mit $\nu_1, \ldots, \nu_r$ bezeichnen.[2] Die entsprechenden theoretischen Wahrscheinlichkeiten für das Auftreten der Elemente in den r Gruppen seien $p_1, \ldots, p_r$, wobei gilt

$$(15.2) \qquad p_i(\theta_1, \ldots, \theta_s) = P(X_i;\ \theta_1, \ldots, \theta_s) \quad \text{für } (i=1,\ldots,r).$$

Für die Häufigkeiten ν_i gilt

$$(15.3) \qquad \sum_{i=1}^{r} \nu_i = n\ ,$$

wobei also n der Stichprobenumfang von X, d.h. die Zahl der Stichprobenelemente von X ist. Die Summe aller Wahrscheinlichkeiten p_i ist 1, d.h.

$$(15.4) \qquad \sum_{i=1}^{r} p_i = 1\ .$$

---

1) Vgl. Cramér: a.a.O., S. 506 u. S. 424-426.

2) Gruppe X_i beispielsweise hat ν_i Gruppenelemente, in Gruppe X_i sind also ν_i Elemente enthalten.

Wenn wir nun die Pearsonsche χ^2-Testformel bilden, erhalten wir

$$(15.5) \qquad \chi^2 = \sum_{i=1}^{r} \frac{(\nu_i - np_i)^2}{np_i} .$$

Hier wird also die Summe der quadratischen Abweichungen der empirischen Häufigkeiten ν_i von den entsprechenden theoretischen Häufigkeiten np_i gebildet, wobei eine Relativierung mit den theoretischen Häufigkeiten erfolgt. Eine Umformung von (15.5) unter Verwendung von (15.3) und (15.4) ergibt[1)]

$$(15.6) \qquad \chi^2 = \sum_{i=1}^{r} \left(\frac{\nu_i^2}{np_i}\right) - n .$$

Da uns die wahren Parameter $\theta_1, \ldots, \theta_s$ in (15.2) unbekannt sind und aus einer gegebenen Stichprobe X beliebig viele Schätzungen der unbekannten Parameter vorgenommen werden können, bietet uns die χ^2-Minimum-Methode folgende Möglichkeiten:

Wir entscheiden uns für diejenigen Schätzungen $\hat{\theta}_1, \ldots, \hat{\theta}_s$, deren Summe der wie in (15.5) relativierten, quadratische Abweichungen der theoretischen Häufigkeiten von den empirischen Häufigkeiten minimal ist.

Wir minimieren also χ^2 bezüglich der θ_j für $j=1,\ldots,s$, wobei wir berücksichtigen, daß p_i eine Funktion der Parameter $\theta_1, \ldots, \theta_s$ ist. [Vgl. (15.2)]. Die 1. Ableitung von (15.5) bezüglich θ_j lautet also wegen (15.2) im Minimum[2)]

---

1) Vgl. Cramér: a.a.O., S. 417.

2) Vgl. auch Menges, G.: Grundriß der Statistik. Teil 1: Theorie. Köln und Opladen 1968, S. 313.

$$(15.7)\quad \frac{\partial \chi^2}{\partial \rho_j} = -2 \sum_{i=1}^{r} \left[\frac{\nu_i - np_i}{p_i} + \frac{(\nu_i - np_i)^2}{2np_i^2}\right] \frac{\partial p_i}{\partial \rho_j} = 0$$

für $(j = 1, \ldots, s)$.

Da für große n der Einfluß des zweiten Ausdrucks in der eckigen Klammer von (15.7) vernachlässigbar gering wird, wurde der komplizierte Ausdruck von (15.7) reduziert auf

$$(15.8)\quad \sum_{i=1}^{r} \frac{\nu_i - np_i}{p_i} \cdot \frac{\partial p_i}{\partial \rho_j} = 0 \,.$$

Die Methode der Schätzung der Parameter mit Hilfe von (15.8) wird auch als modifizierte χ^2-Minimum-Methode bezeichnet. Sie weist gegenüber der ursprünglichen χ^2-Minimum-Methode den Vorteil der erheblich einfacheren Berechnung auf.

Da (15.8) nach Cramér[1] in $\frac{\partial L}{\partial \rho_j}$ überführt werden kann, ergibt die modifizierte χ^2-Minimum-Methode die gleichen Schätzungen wie die Methode der größten Dichte bei vollständiger Information. Cramér[2] beweist auch die Identität der Schätzungen letzterer Methode mit der nicht modifizierten χ^2-Minimum-Methode. In der folgenden Kritik kann daher die χ^2-Minimum-Methode vernachlässigt werden.

---

1) Cramér: a.a.O., S. 426.

2) Cramér: a.a.O., S. 506.

16. Die Methode der Hilfsvariablen[1)]

Wir betrachten eine Strukturgleichung der Form

$$(16.1) \quad y_{1t} = \beta_{12}\, Y_{2t}' + \gamma_1^*\, X_t^{*\prime} + u_{1t} \qquad (t=1,\ldots,T),$$

wobei

$$(16.2) \quad \beta_{12} = [\beta_{12} \ldots \beta_{1G\Delta}], \quad Y_{2t}' = \begin{bmatrix} y_{12} \\ \vdots \\ y_{1G\Delta} \end{bmatrix}, \quad \gamma_1^* = [\gamma_{11} \ldots \gamma_{1K*}]$$

$$X_t^{*\prime} = \begin{bmatrix} x_{11} \\ \vdots \\ x_{1K*} \end{bmatrix}.$$

Aus (16.1) seien K** vorherbestimmte Variablen, nämlich

$$(16.3) \quad X_t^{**\prime} = \begin{bmatrix} x_{1,K*+1} \\ \cdot \\ \cdot \\ \cdot \\ x_{1,K**} \end{bmatrix}$$

ausgeschlossen[2)] durch a-priori-Restriktionen (Nullrestriktionen), so daß gilt

$$(16.4) \quad X_t = [X_t^* \quad X_t^{**}].$$

Analog den "Normalgleichungen" der gewöhnlichen Methode der kleinsten Quadrate erhalten wir dann für

---

1) Vgl. Menges (1961): a.a.O., S. 95-96.

2) Diese Variablen seien jedoch Variablen des Gleichung (16.1) enthaltenden Gleichungssystems.

(16.5) $\hat{u}_{1t} X_t = 0,$

wobei 0 die Nullmatrix der Ordnung (T x K) ist, aus (16.1)

(16.6) $y_{1t} X_t = \hat{\beta}_{12} Y'_{2t} X_t + \hat{\gamma}_1^* X_t^{*\prime} X_t.$

Hieraus erhalten wir für den Fall, daß in (16.1) keine weiteren endogenen Variablen Y'_{2t} auftreten, und sich somit die Verwendung von $X_t^{**\prime}$ erübrigt, also $X_t^{*\prime} = X'_t$, die "Normalgleichung" der gewöhnlichen Methode der kleinsten Quadrate.

(16.7) $y_{1t} X_t = \hat{\gamma}_1^* X'_t X_t.$

Nach Menges[1] besitzt die hier beschriebene Methode der Hilfsvariablen zwar den Vorzug, durch relativ einfache Berechnung konsistente Schätzungen zu erzielen, doch sei die Auswahl der "Hilfsvariablen" X_t^{**} mehr oder weniger willkürlich. Bei der Auswahl der Hilfsvariablen sollte jedoch mindestens folgende Bedingung berücksichtigt werden:

"Man sollte möglichst voneinander unabhängige Hilfsvariablen auswählen, die mit den abhängigen Variablen möglichst stark korreliert sind".[2]

Die Methode der Hilfsvariablen wird im kritischen Teil der Arbeit vernachlässigt, weil sie zu sehr von subjektiven Erwägungen abhängt.

---

1) Menges (1961): a.a.O., S. 96.

2) Menges (1961): a.a.O., S. 96.

17. Sonstige Schätzmethoden

Die dargestellten Methoden sind nur eine Auswahl aus der Menge der bereits entwickelten Schätzmethoden. Gleichwohl wird angenommen, daß diese Auswahl repräsentativ ist für die Untermenge der ökonometrischen Schätzmethoden, d.h. solcher Methoden, die einmal den stochastischen Charakter ökonomischer Funktionen berücksichtigen und die zum anderen zur Untersuchung ökonomischer Phänomene geeignet sind. Daß es sich bei der bisherigen Darstellung nur um eine Auswahl handelt, erkennt man daran, daß solche Methoden, wie etwa die method of mixed estimation von Theil[1)], die Positionsfunktionen von Fisz[2)] und die Verallgemeinerung der Methode der kleinsten Quadrate von

---

1) Vgl. Theil (1961): a.a.O., S. 233-234: Die Methode berücksichtigt explizit die a-priori-Informationen des Ökonometrikers, etwa, daß in einer Funktion das absolute Glied oder ein anderer Parameter aus theoretischen Gründen nicht negativ oder nicht größer als 1 sind oder in einem vorgegebenen Intervall liegen sollen. Diese Informationen werden dann in den Schätzvorgang einbezogen bei der mixed estimation, während sie bei der sonst üblichen Schätzung nur zur Beurteilung der Schätzergebnisse herangezogen werden. Freilich können auch solche Informationen lediglich auf Vermutungen beruhen und dies führt gewissermaßen zu einer Institutionalisierung des Subjektivismus. Das Schätzergebnis würde dann präjudiziert, was bedenklich erscheint.

2) Vgl. Fisz, M.: Wahrscheinlichkeitsrechnung und mathematische Statistik. 4. Aufl., Berlin 1966, S. 308-310. Vgl. hierzu auch Menges (1968): a.a.O., S. 319-321. Diese Methode wurde nicht behandelt, da ein Vergleich der Lageparameter mit den Schätzparametern der behandelten Methode im folgenden kritischen Teil insbesondere beim ökonometrischen Modell keine allgemeingültigen Ergebnisse erzielt.

Chow[1] nicht behandelt wurden. Im übrigen werden im folgenden kritischen Teil nur diejenigen ökonometrischen Schätzmethoden einem kritischen Vergleich unterzogen, die sich einmal in wesentlichen Punkten voneinander unterscheiden, die zum anderen in ihrem Aufbau nicht von subjektiven Eingriffen des Ökonometrikers abhängig sind und die schließlich nicht nur für spezielle, engdefinierte Probleme konstruiert sind. Diese Restriktionen sind erforderlich, um durch den Vergleich mehrerer Schätzverfahren miteinander auf der objektiven Basis ökonometrischer Hauptanwendungsfälle zu nachvollziehbaren Resultaten zu gelangen.

---

1) Chow, G.C.: A comparison of alternative estimators for simultaneous equations. Econometrica, 32(1964), S. 532-553. Diese Verallgemeinerung ist im Grunde kein neues Schätzverfahren sondern nur eine Orientierung der Methode der größten Dichte in Richtung auf die Methode der kleinsten Quadrate. In der Tat ist diese verallgemeinerte Methode identisch mit der Methode der größten Dichte und wird somit doch noch im folgenden kritischen Teil implizit mitbehandelt, wenn von der Methode der größten Dichte die Rede ist.

Z W E I T E R T E I L

Kritischer Vergleich

18. Abgrenzung

Die kritische Betrachtung der wichtigsten der bisher dargestellten ökonometrischen Schätzverfahren wird zunächst die Auswirkungen der bei der Darstellung zum Vorschein gekommenen Unterschiede der einzelnen Schätzverfahren auf die praktische Anwendbarkeit verfolgen müssen. Hierzu ist es erforderlich, verschiedene praktische Anwendungsfälle voneinander abzugrenzen.

(1) Die wohl einfachste Abgrenzung ist die nach der Zahl der zu schätzenden Strukturgleichungen eines Gleichungssystems.

Hier unterscheiden wir

(a) Gleichungssysteme mit einer Strukturgleichung und
(b) Gleichungssysteme mit mehr als einer Strukturgleichung.

(2) Gleichungssysteme mit mehr als einer Strukturgleichung können ihrerseits wieder unterschieden werden in[1)]

(a) rekursive Gleichungssysteme und
(b) interdependente Gleichungssysteme.

Rekursiv nennt man solche Gleichungssysteme , in denen die endogenen Variablen des Gleichungssystems nicht gemeinsam abhängig sind. Zwei endogene Variablen sind gemeinsam abhängig, wenn sie in zwei Gleichungen jeweils zusammen auftreten. Tritt dagegen nur eine der endogenen Variablen in beiden Gleichungen zugleich auf, die andere hingegen nur in z.B. der ersten Gleichung, so lassen sich die beiden Gleichungen sukzessiv auflösen durch Auflösung

---

1) Wold, H. (in Zusammenarbeit mit L. Juréen): a.a.O., S. 48-53, s. bes. S. 50.

der zweiten und dann der ersten Gleichung. Somit erkennt man rekursive Gleichungssysteme an der Dreiecksform (Triangularität) der Matrix der Parameter der endogenen Variablen. Entsprechend kann Wold[1] aus dieser Triangularität kausale Abhängigkeiten und Kausalketten herleiten. Dagegen treten in interdependenten Gleichungssystemen mindestens zwei der endogenen Variablen gleichzeitig jeweils in mindestens zwei verschiedenen Gleichungen des Systems auf, d.h. diese endogenen Variablen sind also gemeinsam abhängig. Daher nennt man die endogenen Variablen der betroffenen Gleichungen des interdependenten Gleichungssystems auch gemeinsam abhängige Variablen.

(3) Nach der Zusammensetzung der vorherbestimmten Variablen lassen sich schließlich noch unterscheiden

(a) Gleichungssysteme, die keine verzögerten endogenen Variablen enthalten.
(b) Gleichungssysteme, die verzögerte endogene Variablen enthalten.

Diese letzte Unterscheidung ist natürlich auch für Gleichungssysteme mit nur einer Strukturgleichung von großer Bedeutung.

(4) Schließlich sind noch zu unterscheiden wie bereits im ersten Teil[2] dieser Arbeit

(a) das System der Strukturgleichungen und
(b) das System von reduzierte-Form-Gleichungen, welches aus dem System der Strukturgleichung durch Auflösung nach den endogenen Variablen hervorgeht.

1) Wold, H. (in Zusammenarbeit mit L. Juréen): a.a.O., S. 50, vgl. auch Wold H.: A generalization of causal chain models. Econometrica, 28(1960), S. 443-463. Vgl. weiter Wold, H.: Forecasting by the chain principle. In: Econometric Model Building. Hrsg.: H.O.A. Wold. Amsterdam 1964, S. 5-36, s. bes. S. 18-19. Siehe auch Strotz, R.H. u. Wold, H.: Recursive versus nonrecursive systems: An attempt at synthesis. Econometrica, 28(1960), S. 417-427, s. bes. S. 417.
2) Vgl. im Kapitel 9 Gleichung (9.1) bzw. (9.5).

19. Die Eigenschaften guter Schätzungen[1)]

1. Erwartungstreu oder unverzerrt ist eine Schätzung $\hat{\delta}$, wenn ihr Erwartungswert gleich dem "wahren" Wert, δ, ist:

$$(19.1) \qquad E(\hat{\delta}) = \delta.$$

Weicht der Erwartungswert der Schätzung $\hat{\delta}$ von ihrem "wahren" Wert ab, so bezeichnet man dies als Verzerrung[2)]

$$(19.2) \qquad d = E(\hat{\delta}) - \delta.$$

2. Effizient ist eine erwartungstreue Schätzung $\hat{\delta}$, wenn ihre Streuung um den "wahren" Wert δ kleiner/gleich der Streuung jeder anderen Schätzung $\hat{\delta}^*$ um den "wahren" Wert δ ist:

$$(19.3) \qquad E(\hat{\delta} - \delta)^2 \leqq E(\hat{\delta}^* - \delta)^2.$$

Die Effizienz kann als eine relative (im Vergleich zu einer anderen Schätzmethode) Eigenschaft einer Schätzmethode betrachtet werden. Eine Schätzung ist dagegen absolut effizient, wenn sie die minimale Streuung um den "wahren" Parameter aufweist im Vergleich zu allen anderen Schätzungen, die die gleichen Informationen verwenden.

3. Konsistent ist eine Schätzung $\hat{\delta}$, wenn mit zunehmendem Stichprobenumfang T die Wahrscheinlichkeit, daß $\hat{\delta}$ von dem "wahren" Wert δ um weniger als einen vorgegebenen, beliebig kleinen, absoluten Betrag ε abweicht, gegen den Wert Eins strebt:

---

1) Vgl. Menges (1961): a.a.O., S. 88-90. Die dort aufgeführten Eigenschaften werden hier in gekürzter Form wiederholt, da sie im folgenden immer wieder herangezogen werden.

2) Vgl. auch Kendall, M.G. u. Stuart, A.: The Advanced Theory of Statistics. Bd. 2: Inference and Relationship. London 1961, S. 6.

$$(19.4) \qquad \lim_{T\to\infty} P(|\hat{\delta} - \delta| < \varepsilon) = 1.$$

Diese Grenzwertbetrachtung kann auch durch den Wahrscheinlichkeitslimes plim ausgedrückt werden:

$$(19.5) \qquad \operatorname{plim}_{T\to\infty} \hat{\delta} = \delta.$$

"Die Konsistenz ist die wichtigste Eigenschaft einer Schätzung, sie ist gewissermaßen die Minimalforderung, die man an die Güte einer Schätzung stellt."[1)]

Die Eigenschaften[2)] "asymptotische Normalität" und "asymptotische Effizienz" als asymptotische Eigenschaften wurden in dieser Arbeit im Rahmen des kritischen Vergleichs nicht näher untersucht.

---

1) Menges (1961): a.a.O., S. 89.
2) Vgl. Menges (1961): a.a.O., S. 89-90.

20. Eine einzige Gleichung mit einer endogenen Variablen

Die Gleichung laute[1]

$$(20.1) \quad y_t = \pi_1 z_{1t} + \pi_2 z_{2t} + \ldots + \pi_K z_{Kt} + v_t$$

$$(t=1,\ldots,T)$$

und sei identifizierbar[2].

Nach unserer Abgrenzung (3)[3] bestehen für eine einzige Gleichung die beiden Möglichkeiten,

a) daß die Gleichung keine verzögerten endogenen Variablen enthält,
b) daß verzögerte endogene Variablen innerhalb der Gruppe der vorherbestimmten Variablen auftreten.

Koopmans und Hood[4] haben diese Einteilung noch weiter aufgegliedert, indem sie für die gemeinsame Verteilungsfunktion der latenten Variablen eine weitere Unterscheidung berücksichtigten

---

1) Koopmans u. Hood: a.a.O., S. 132.

2) Da die Identifizierbarkeit als die Grundlage der Schätzbarkeit schlechthin angesehen werden kann, wird im folgenden immer implizit vorausgesetzt, daß die angeführten Gleichungen identifizierbar sind. Eine ähnliche implizite und fundamentale Voraussetzung ist, daß Beobachtungswerte für alle beobachtbaren Variablen (d.h. alle Variablen außer der latenten Variablen) für alle betrachteten Perioden überhaupt zur Verfügung stehen. Auch auf die Wiederholung dieser conditio sine qua non möge im folgenden verzichtet werden.

3) Vgl. Kapitel 19.

4) Vgl. Koopmans u. Hood: a.a.O., S. 131-133.

α) "Die gemeinsame Verteilung der v_t ist das Ergebnis unabhängiger und identischer Normalverteilungen aller v_t $(t=1,\ldots,T)$ mit Mittelwert $E(v_t) = 0$ und Streuung $E(v_t^2) = \omega$"[1)2)]

ß) "Die gemeinsame Verteilung der v_t ist nicht als normal spezifiziert; die v_t sind unabhängig und identisch verteilte Zufallsvariablen mit Mittelwert $E(v_t) = 0$ und endlicher Streuung $E(v_t^2) = \omega$ für $t=1,\ldots,T$."[1)2)]

Hierbei soll ß) eine Verallgemeinerung von α) sein. Durch die Anwendung der beiden Einteilungen a oder b und α oder ß gelangen Koopmans und Hood zu 4 verschiedenen Möglichkeiten, die der Anwendung eines Schätzverfahrens zugrundeliegen können.

Für die gewöhnliche Methode der kleinsten Quadrate gelangen Koopmans und Hood bei Vorliegen der alternativen Voraussetzungen zu folgenden Konsequenzen für die Schätzungen:[3)]

1) Bei einem Zusammentreffen der Bedingungen a und α sind die Kleinst-Quadrat-Schätzungen $\hat{\pi}_1, \hat{\pi}_2, \ldots, \hat{\pi}_K$ der Parameter $\pi_1, \pi_2, \ldots, \pi_K$ konsistent, erwartungstreu und effizient und demnach auch beste lineare unverzerrte Schätzungen.

2) Sind die Bedingungen a und ß erfüllt, so sind die Kleinst-Quadrat-Schätzungen konsistent und, da das Markoff-Theorem[4)] für Kleinst-Quadrat-Schätzungen gültig bleibt, ebenfalls beste lineare unverzerrte Schätzungen.

---

1) Koopmans u. Hood: a.a.O., S. 132.

2) Wird in (20.1) v_t als Spaltenvektor der Länge T aufgefaßt, so ist die letztere Bedingung so zu formulieren: "und Streuungs-Kovarianz-Matrix $V = E(v_t v_t') = \sigma^2 I$."

3) Vgl. Koopmans u. Hood: a.a.O., S. 133.

4) Vgl. Menges (1961): a.a.O., S. 94.

3) Liegen die Bedingungen b und α vor, so sind die Kleinst-Quadrat-Schätzungen konsistent und noch asymptotisch effizient.

4) Wenn die Bedingungen b und ß vorliegen, so sind die Kleinst-Quadrat-Schätzungen lediglich konsistent.

Mit der gewöhnlichen Methode der kleinsten Quadrate erzielte Schätzungen der Parameter eines nur aus einer Gleichung bestehenden vollständigen Gleichungssystems sind also immer konsistent, solange die latenten Variablen v_t unabhängig und identisch verteilte Zufallsvariablen mit Mittelwert $E(v_t) = 0$ und endlicher Streuung $E(v_t^2) = \omega$ für $t=1,\ldots,T$ sind. Die gemeinsame Verteilung der v_t muß hierbei nicht einmal als Normalverteilung spezifiziert sein.

Die Konsequenzen 1) bis 4) gelten unter gegebenen Bedingungspaaren ebenfalls für die indirekte Methode der kleinsten Quadrate, da Gleichung (20.1) nur eine unverzögerte endogene Variable, nämlich y_t enthält. Daher ist Gleichung (20.1) eine reduzierte-Form-Gleichung, weil keine weiteren unverzögerten endogenen Variablen auftreten.

Da somit die Berechnung der reduzierten Form und die Umwandlung der Parameterschätzungen der reduzierten Form in Parameterschätzungen der Strukturgleichung nicht mehr erforderlich sind, führen unter den gegebenen Bedingungen die direkte - d.h. die gewöhnliche - und die indirekte Methode der kleinsten Quadrate zu identischen Schätzwerten der Parameter der Strukturgleichung und damit zu identischen Eigenschaften[1] der Schätzungen.

Wie wir bereits im 4. Kapitel[2] gesehen haben, sind die verallgemeinerte Methode der kleinsten Quadrate und die

1) Vgl. Koopmans u. Hood: a.a.O., S. 133.
2) Vgl. die Betrachtungen im Anschluß an Gleichung (4.24).

gewöhnliche Methode der kleinsten Quadrate identisch, wenn Homoskedastizität vorliegt und zugleich die latenten Variablen nicht autokorreliert sind. Nur in diesem Spezialfall hat die Streuungs-Kovarianz-Matrix der latenten Variablen das Aussehen einer Skalarmatrix[1]).

Hieraus erkennen wir, daß die Konsequenzen 1) bis 4), die unter den gegebenen Bedingungen für die gewöhnliche Methode der kleinsten Quadrate gelten, auch für die verallgemeinerte Methode der kleinsten Quadrate zutreffen, wenn die Streuungs-Kovarianz-Matrix der latenten Variablen eine Skalarmatrix ist, wenn also die Kovarianzen der latenten Variablen der Gleichung (20.1) für alle Perioden $t = 1, \ldots, T$ gleich Null sind und zugleich die Streuungen der latenten Variablen in verschiedenen Perioden einander gleich, also zeitlich invariant sind.

Für die zweistufige Methode der kleinsten Quadrate ergeben sich für eine einzige Gleichung der Art (20.1) die gleichen Konsequenzen 1) bis 4) wie für die gewöhnliche Methode der kleinsten Quadrate unter den gegebenen Bedingungen.

Zur besseren Vergleichbarkeit schreiben wir Gleichung (20.1) in Matrizen

$$(20.2) \qquad y_t = \Pi\, z_t' + v_t \qquad (t=1,\ldots,T),$$

---

1) Vgl. Zurmühl: a.a.O., S. 12. Eine Skalarmatrix kann als das Produkt eines Skalars mit einer Einheitsmatrix betrachtet werden. Daher ist die Inverse einer Skalarmatrix diejenige Matrix, die durch die Multiplikation des Kehrwertes dieses Skalars mit der Einheitsmatrix der Ordnung der Skalarmatrix entsteht. Da Skalare bei der Multiplikation vorgezogen werden dürfen, kürzt sich der Skalar gegen seinen Kehrwert in Gleichung (4.24), wenn V eine Skalarmatrix ist und wir erhalten (2.29), die Parametervektorschätzung nach der gewöhnlichen Methode der kleinsten Quadrate.

wobei y_t und v_t Skalare sind und

$$(20.3) \qquad \Pi = [\Pi_1 \ \ldots \ \Pi_K] \quad \text{und} \quad z_t' = \begin{bmatrix} z_{1t} \\ \vdots \\ z_{Kt} \end{bmatrix} .$$

Wir sehen sofort[1], daß in Gleichung (20.2) keine anderen unverzögerten endogenen Variablen vorkommen als y_t, so daß der 1. Schritt der zweistufigen Methode der kleinsten Quadrate, nämlich die Schätzung $\hat{Y}_2$ der anderen unverzögerten endogenen Variablen Y_2 aufgrund der vorherbestimmten Variablen, entfällt. In dieser einfachen Form (20.2) kann also nur noch der 2. Schritt der zweistufigen Methode der kleinsten Quadrate, d.h. die Schätzung von (20.2) mit der gewöhnlichen Methode der kleinsten Quadrate, angewandt werden, so daß die resultierenden Schätzwerte offensichtlich die Schätzwerte der gewöhnlichen Methode der kleinsten Quadrate sind. Wir sehen somit, daß die zweistufige Methode der kleinsten Quadrate mit den Ergebnissen der gewöhnlichen Methode der kleinsten Quadrate übereinstimmende Ergebnisse erzielt, solange in der zu schätzenden Gleichung nur eine einzige unverzögerte endogene Variable enthalten ist.

Wenn Gleichung (20.1) die einzige Gleichung eines Gleichungssystems ist, stimmen die dreistufige Methode der kleinsten Quadrate und die gewöhnliche Methode der kleinsten Quadrate überein. Die erste Stufe der dreistufigen Methode entfällt - wie bei der zweistufigen Methode - weil nur eine unverzögerte endogene Variable in (20.1) auftritt und somit keine weiteren unverzögerten endogenen Variablen aufgrund der vorherbestimmten Variablen neugeschätzt werden müssen. Die dritte Stufe der dreistufigen Methode entfällt, weil das Gleichungssystem bereits durch die Gleichung

1) Vgl. mit den Gleichungen (5.1) und (5.19).

(20.1) vollständig beschrieben ist, da keine weiteren Gleichungen existieren. Da die zweite Stufe der dreistufigen Methode die Schätzung von (20.1) mit der gewöhnlichen Methode der kleinsten Quadrate ist, gelten in diesem Fall die Konsequenzen 1) bis 4) der gewöhnlichen Methode der kleinsten Quadrate auch für die dreistufige Methode der kleinsten Quadrate.

Ist Gleichung (20.1) nur eine - etwa die erste - Gleichung eines größeren Gleichungssystems, so stimmt die dreistufige Methode der kleinsten Quadrate mit der gewöhnlichen Methode der kleinsten Quadrate überein, wenn die Streuungs-Kovarianz-Matrix der latenten Variablen aller Gleichungen des größeren, vollständigen Gleichungssystems eine Diagonalmatrix ist, wobei mindestens die Streuungsmatrix der latenten Variablen der Gleichung (20.1) - dies ist für die erste Gleichung die Untermatrix V_{11} in (6.20) - eine Skalarmatrix sein muß. In diesem Fall treffen die Konsequenzen 1) bis 4) für die dreistufige Methode der kleinsten Quadrate nur zu, wenn also folgende zusätzlichen Bedingungen erfüllt sind:

Einmal müssen alle Kovarianz-Matrizen nur aus Nullen bestehen (Nicht nur die Autokorrelation ist ausgeschlossen, auch die gleichzeitigen Kovarianzen der latenten Variablen verschiedener Strukturgleichungen, die meist als zeitlich invariant angenommen werden, müssen jetzt verschwinden.).

Zum anderen dürfen die latenten Variablen der Gleichung (20.1) nicht autokorreliert und ihre Streuung muß zeitlich invariant sein.

Interessieren uns die anderen Gleichungen, außer (20.1), des größeren Gleichungssystems nicht, so ist die Form ihrer Streuungs-Matrizen der latenten Variablen keinen Restriktionen unterworfen, und wir müssen nicht mehr fordern, daß in

(6.20) und (6.21) V eine Diagonalmatrix ist. Gleichwohl muß dann V zumindest block-diagonal sein, da die Kovarianz-Matrizen in jedem Fall Null-Matrizen sein müssen, um zumindest für Gleichung (20.1) eine Identität der Schätzungen nach der gewöhnlichen Methode der kleinsten Quadrate mit denen der dreistufigen Methode der kleinsten Quadrate zu erzielen.

Für die Maximum-Likelihood-Methode bei vollständiger Information können nur die Konsequenzen 1) und 3) Gültigkeit besitzen, da eine der wesentlichsten Annahmen der Maximum-Likelihood-Methode die Annahme normalverteilter latenter Variablen ist. Die Gültigkeit der Konsequenz 3), die bei Vorhandensein verzögerter endogener Variablen eintritt, bewiesen Koopmans und Hood[1], indem sie, von den Bedingungen α und b ausgehend, die Parameter einer einzigen Gleichung mit der Maximum-Likelihood-Methode schätzten und die Schätzformel für die gewöhnliche Methode der kleinsten Quadrate erhielten.

Da sogar mit der strengen Bedingung b eine Identität der Parameterschätzungen der Maximum-Likelihood-Methode mit den Parameterschätzungen der Methode der kleinsten Quadrate nachgewiesen wurde, konnten Koopmans und Hood[2] folgern, daß durch die Verbindung der erleichternden Bedingungen a mit der Normalverteilungsbedingung α erst recht eine Identität der Parameterschätzungen beider Methoden erzielt wird. Damit wurde auch die Gültigkeit der Konsequenz 1) für die Maximum-Likelihood-Methode bei vollständiger Information bewiesen.

---

1) Vgl. Koopmans u. Hood: a.a.O., S. 147-150.
2) Vgl. Koopmans u. Hood: a.a.O., S. 150.

Für die Maximum-Likelihood-Methode bei beschränkter Information haben ebenfalls die Konsequenzen 1) und 3) Gültigkeit, wenn man bedenkt, daß bei einem nur aus einer einzigen Gleichung bestehenden vollständigen Gleichungssystem die Information nicht durch Vernachlässigung von anderen Gleichungen beschränkt werden kann. Wie für die Maximum-Likelihood-Methode bei vollständiger Information können hier die Konsequenzen 2) und 4) keine Gültigkeit besitzen, da die Normalverteilungsannahme auch für die Maximum-Likelihood-Methode bei beschränkter Information wesentlich ist.

21. Ein reduzierte-Form-Gleichungssystem[1], das keine verzögerten endogenen Variablen enthält

Wir betrachten nach dem einfachen Fall einer einzelnen Gleichung den etwas schwierigeren Fall eines Gleichungssystems, das aus G Gleichungen besteht, die in der reduzierten Form geschrieben sind. Somit tritt in jeder Gleichung nur eine der endogenen Variablen des Gleichungssystems auf. Wir betrachten vorerst nur den Fall, bei dem keine verzögerten endogenen Variablen auftreten, die vorherbestimmten Variablen also nur aus exogenen Variablen bestehen.

Unser reduzierte-Form-Gleichungssystem habe folgendes Aussehen

$$(21.1) \qquad y_t' = -\Pi\, z_t' + v_t' \qquad (t=1,\ldots,T),$$

wobei

$$(21.2) \quad y_t' = \begin{bmatrix} y_{1t} \\ \cdot \\ \cdot \\ \cdot \\ y_{Gt} \end{bmatrix}, \quad \Pi = \begin{bmatrix} \Pi_{11} & \cdots & \Pi_{1K} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \Pi_{G1} & \cdots & \Pi_{GK} \end{bmatrix}, \quad z_t' = \begin{bmatrix} z_{1t} \\ \cdot \\ \cdot \\ \cdot \\ z_{Kt} \end{bmatrix} \text{ und } v_t' = \begin{bmatrix} v_{1t} \\ \cdot \\ \cdot \\ \cdot \\ v_{Gt} \end{bmatrix}.$$

z_t' soll hier also nur exogene Variablen enthalten.

Zunächst betrachten wir die Schätzungen nach der Methode der größten Dichte bei vollständiger Information.

Hierzu formen wir (21.1) um zu

$$(21.3) \qquad Ax_t' = y_t' + \Pi z_t' = v_t' \qquad (t=1,\ldots,T),$$

---

1) Vgl. Koopmans u. Hood: a.a.O., S. 151-155.

wobei

$$(21.4) \qquad A = [I_G \quad \Pi] \text{ und } x_t' = \begin{bmatrix} y_t' \\ z_t' \end{bmatrix}.$$

Wir verwenden (21.3) und (21.4) bei der Formulierung der auf A konzentrierten logarithmischen Likelihoodfunktion vom Typ (8.23) und erhalten

$$(21.5) \qquad L(A) = -\frac{G}{2}(1 + \ln 2\pi) - \frac{1}{2}\ln(\det A\, M_{xx}\, A'),$$

weil $\ln |\det I_G| = 0$, wobei 0 der Skalar Null ist.

Die Maximierung von (21.5) bezüglich A ergibt den komplizierten Ausdruck

$$(21.6) \qquad \frac{\partial L(A)}{\partial A} = -\,Sp\,\{[A\, M_{xx}\, A']^{-1}\, A\, M_{xx}\} = 0.$$

Zu einem leichter vergleichbaren Ergebnis gelangt man, wenn man die zusätzliche Restriktion trifft, daß die Streuungs-Kovarianz-Matrix der latenten Variablen eine Diagonalmatrix sein soll[1]. Analog zu (8.31) erhalten wir dann die für jede reduzierte-Form-Gleichung K Normalgleichungen

$$(21.7) \qquad \frac{1}{T}\sum_{t=1}^{T} v_{it}\, x_j = 0, \qquad (i=1,\ldots,G),\ (j=1,\ldots,K),$$

weil die Differentiation $\frac{\partial I_G}{\partial \hat{a}_{ij}}$ Null ergibt.

---

1) Vgl. (8.26) und die anschließenden Berechnungen.

Aus den Normalgleichungen (21.7) erhalten wir die Schätzungen nach der gewöhnlichen Methode der kleinsten Quadrate.

Somit führen die Schätzungen nach der Methode der größten Dichte bei vollständiger Information zu identischen Schätzungen wie die Methode der kleinsten Quadrate, wenn

a) das Gleichungssystem in reduzierter Form geschrieben ist (die reduzierte Form somit existiert),

b) die latenten Variablen normal verteilt sind,

c) die latenten Variablen stochastisch unabhängig voneinander sind und einer gemeinsamen Dichtefunktion gehorchen und

d) die vorherbestimmten Variablen keine verzögerten endogenen Variablen enthalten und von den latenten Variablen stochastisch unabhängig sind.

Unter diesen Bedingungen führen beide Methoden zu konsistenten Schätzungen.[1)]

Wir betrachten nun die Schätzungen mit der Methode der größten Dichte bei beschränkter Information.

Wiederum soll das Gleichungssystem in reduzierter Form geschrieben sein und keine verzögerten endogenen Variablen enthalten.

---

1) Vgl. die Bemerkungen am Ende des 8. Kapitels.

Die logarithmische Likelihoodfunktion nach der Methode der größten Dichte bei beschränkter Information lautet[1)]

$$(21.8)\quad L(A_I) = -\frac{G}{2}(1 + \ln 2\pi) - \frac{1}{2}\ln\det(A_I M_{xx} A_I') + \frac{1}{2}\ln\det(B_I W B_I') - \frac{1}{2}\ln\det W.$$

Da unser Gleichungssystem in der reduzierten Form geschrieben ist, ist B die Einheitsmatrix. Demnach ist auch die Untermatrix B_I eine Einheitsmatrix und die letzten beiden Ausdrücke von (21.8) heben sich gegeneinander auf. Wir erhalten also

$$(21.9)\quad L(A_I)_R = -\frac{G}{2}(1 + \ln 2\pi) - \frac{1}{2}\ln\det(A_I M_{xx} A_I').$$

Da nur noch in dem letzten Ausdruck die Parameter A_I vorkommen, wird die logarithmische Likelihoodfunktion (21.9) maximiert, wenn der letzte, negative Ausdruck minimiert wird. Die Analogie zur Methode der kleinsten Quadrate wird hier schon offenkundig.

Die Differentiation von (21.9) bezüglich A_I ergibt

$$(21.10)\quad \frac{\partial L(A_I)_R}{\partial A_I} = -\frac{1}{2}\,\frac{\partial \ln\det(A_I M_{xx} A_I')}{\partial A_I}$$

Wir müssen nun zwei Fälle unterscheiden: Besteht das Gleichungssystem I nur aus einer einzigen Gleichung, so ist $A_I M_{xx} A_I'$ zweifellos ein Skalarprodukt, also ein

---

1) Vgl. Gleichung (10.48).

Skalar, dessen Determinante der Skalar selbst ist und dessen logarithmus naturalis die Maximierung nicht beeinflußt. Dann erhalten wir das Maximum durch die Differentiation

$$(21.11)\quad \frac{\partial\, A_I\, M_{xx}\, A_I'}{\partial\, A_I} = \frac{\partial(M_{yy} + \Gamma_I\, M_{zy} + M_{yz}\, \Gamma_I' + \Gamma_I M_{zz} \Gamma_I')}{\partial\, \Gamma_I} = 0,$$

welche sich unter Berücksichtigung der Aufteilung von A_I und M_{xx} ergibt:

$$(21.12)\quad A_I = [1 \quad \Gamma_I] \quad \text{und} \quad M_{xx} = \begin{bmatrix} M_{yy} & M_{yz} \\ M_{zy} & M_{zz} \end{bmatrix}.$$

Aus (21.11) erhalten wir dann, da die Skalarprodukte $\Gamma_I\, M_{zy}$ und $M_{yz}\, \Gamma_I'$ identisch sind

$$(21.13)\quad M_{yz} + \Gamma_I\, M_{zz} = 0,$$

woraus sich die Kleinst-Quadrat-Schätzung

$$(21.14)\quad \Gamma_I = - M_{yz}\, M_{zz}^{-1}$$

ergibt.

Besteht das Gleichungssystem I jedoch aus mehr als einer Gleichung, so läßt sich nur für den Diagonalfall analog zu dem Vorgehen zur Gewinnung von Gleichung (21.7) eine Identität mit den Schätzungen nach der Methode der kleinsten Quadrate herleiten. Hierin kommt wieder die Bedeutung

der Interdependenzen zwischen den einzelnen reduzierte-Form-Gleichungen zum Ausdruck, die sich in der Streuungs-Kovarianz-Matrix ihrer latenten Variablen niederschlagen.

Damit zeigt sich, daß für das Teilsystem (oder die einzelne Gleichung) I die Schätzungen nach der Methode der größten Dichte und die Schätzungen nach der Methode der kleinsten Quadrate identisch sind, wenn

a. das betrachtete Gleichungssystem in reduzierter Form geschrieben ist,

b. die latenten Variablen normalverteilt sind,

c. die latenten Variablen stochastisch unabhängig voneinander sind und einer gemeinsamen Dichtefunktion gehorchen,

d. die vorherbestimmten Variablen keine verzögerten endogenen Variablen enthalten und von den latenten Variablen stochastisch unabhängig sind und

e. nur wenn das Teilsystem I aus mehr als einer Gleichung besteht: die Streuungs-Kovarianz-Matrix der latenten Variablen eine Diagonalmatrix ist.

Betrachten wir nun die Schätzung mit der verallgemeinerten Methode der kleinsten Quadrate.

Diese ist[1] bei einem Gleichungssystem mit der Schätzung nach der gewöhnlichen Methode der kleinsten Quadrate dann und nur dann identisch, wenn die Streuungs-Kovarianz-Matrix der latenten Variablen eine Diagonalmatrix ist und zugleich Homoskedastizität vorliegt. Nur in diesem Fall läßt sich die Streuungs-Kovarianz-Matrix der latenten

1) Vgl. die Betrachtungen im Anschluß an Gleichung (4.24).

Variablen V in der Form $\sigma^2(I)$ schreiben, wobei σ^2 ein Skalar und I die Einheitsmatrix ist, so daß nach Kürzung mit $\frac{1}{\sigma^2}$ die Schätzung nach der gewöhnlichen Methode der kleinsten Quadrate resultiert.

Wenn also beispielsweise Autokorrelation der latenten Variablen vorliegt, kann von einer Identität der Schätzungen der verallgemeinerten Methode der kleinsten Quadrate mit den Schätzungen der gewöhnlichen Methode der kleinsten Quadrate nicht mehr die Rede sein. Autokorrelation[1] kann durch fehlerhafte Spezifikation der Variablen, fehlerhafte Wahl der Funktionsform der Schätzfunktion, Beobachtungsfehler und andere Ursachen bedingt sein. Daher kann ein in reduzierter Form geschriebenes Gleichungssystem, welches keine verzögerten endogenen Variablen enthält, durchaus Autokorrelation aufweisen. Es ergibt sich also eine Diskrepanz zwischen den Schätzungen nach der verallgemeinerten Methode der kleinsten Quadrate und den Schätzungen nach der gewöhnlichen Methode der kleinsten Quadrate, wenn die Streuungs-Kovarianz-Matrix der latenten Variablen nicht eine Diagonalmatrix ist, deren Elemente auf der Hauptdiagonalen einander gleich sind.

Die Schätzungen mit der <u>indirekten Methode der kleinsten Quadrate</u> sind bei einem in reduzierter Form geschriebenen Gleichungssystem identisch mit den aufgrund der gewöhnlichen Methode der kleinsten Quadrate erzielten Schätzungen dieser reduzierte-Form-Gleichungen. In der Tat unterscheidet sich die indirekte Methode der kleinsten Quadrate von der gewöhnlichen Methode der kleinsten Quadrate, welche die Parameter der Strukturgleichungen schätzt, lediglich durch

---

1) Vgl. Schneeweiß: a.a.O., S. 77-78.

1. die Umbildung der Strukturgleichungen in reduzierte-Form-Gleichungen und

2. die Berechnung der Strukturparameterschätzungen auf der Grundlage der zuvor geschätzten reduzierte-Form-Parameter.

Diese beiden Schritte entfallen jedoch, wie in diesem Fall, wenn das Gleichungssystem von Anfang an in der reduzierten Form geschrieben ist, d.h. wenn wir unter Umständen die Strukturgleichungen überhaupt nicht kennen.

Auch die Schätzungen mit der zweistufigen Methode der kleinsten Quadrate sind identisch mit den Schätzungen der gewöhnlichen Methode der kleinsten Quadrate, wenn das Gleichungssystem in der reduzierten Form geschrieben ist. Dies folgt aus der Tatsache, daß in der reduzierten Form in jeder Gleichung nur eine einzige unverzögerte endogene Variable auftritt, so daß die erste Stufe der zweistufigen Methode, nämlich die Schätzung der anderen unverzögerten endogenen Variablen der Gleichung, entfällt. Damit ist nur noch die zweite Stufe der zweistufigen Methode der kleinsten Quadrate relevant; diese zweite Stufe ist jedoch identisch mit der gewöhnlichen Methode der kleinsten Quadrate, so daß die zweistufige Methode der kleinsten Quadrate in diesem Falle keine besseren Ergebnisse erzielen kann als die gewöhnliche Methode der kleinsten Quadrate.

Die Schätzungen mit der dreistufigen Methode der kleinsten Quadrate stimmen nur unter sehr restriktiven Bedingungen mit den Schätzungen nach der gewöhnlichen Methode der kleinsten Quadrate überein. Da die dreistufige Methode eine Gleichungssystem-Schätzmethode ist, kann eine

Übereinstimmung mit der gewöhnlichen Methode der kleinsten Quadrate, die im Grunde eine Ein-Gleichung-Schätzmethode ist, nur dann gewährleistet sein, wenn alle Gleichungen des Gleichungssystems voneinander unabhängig sind; in einem solchen Falle würde man zu denselben Schätzungen gelangen, wenn man entweder

1. mit der dreistufigen Methode der kleinsten Quadrate das gesamte Gleichungssystem oder

2. mit der zweistufigen Methode der kleinsten Quadrate jede Gleichung des Gleichungssystems getrennt oder

3. mit der gewöhnlichen Methode der kleinsten Quadrate jede Gleichung des Gleichungssystems, das jedoch unbedingt ein reduzierte-Form-Gleichungssystem sein muß[1], schätzen würde.

Diese Unabhängigkeit der Gleichungen voneinander ist jedoch nicht schon dadurch gesichert, daß das Gleichungssystem in reduzierter Form geschrieben ist. Das Auftreten von Autokorrelation der latenten Variablen bedeutet auch hier die Aufhebung der Unabhängigkeit der Gleichungen und zwar der aus einer Gleichung durch T Beobachtungsperioden gebildeten T Gleichungen. Da wir außerdem jetzt ein Gleichungssystem von G reduzierte-Form-Gleichungen betrachten, müssen auch die latenten Variablen verschiedener Gleichungen bei gleicher oder verschiedener Periode voneinander unabhängig sein. Unabhängigkeit in diesem strengeren Sinne bedeutet also, daß die Streuungs-Kovarianz-

1) weil in jeder Gleichung dann nur eine einzige unverzögerte endogene Variable auftritt, so daß die Schätzung mit der zweistufigen Methode der kleinsten Quadrate dann der Schätzung mit der gewöhnlichen Methode der kleinsten Quadrate entspricht.

Matrix, V, der latenten Variablen des reduzierte-Form-Gleichungssystems eine Skalarmatrix folgenden Typs sein muß

$$(21.15) \qquad V = \begin{matrix} V_{11} & 0 & \cdots & 0 \\ 0 & V_{11} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & V_{11} \end{matrix} ,$$

wobei gilt $V_{11} = \sigma_{11}\, I$, das heißt σ_{11} ist ein konstanter Skalar und I ist die Einheitsmatrix.[1)]

Die Schätzungen nach der dreistufigen Methode der kleinsten Quadrate stimmen also mit den Schätzungen nach der gewöhnlichen Methode der kleinsten Quadrate vollständig überein, wenn die Streuungs-Kovarianz-Matrix, V, der latenten Variablen des Gleichungssystems von G reduzierte-Form-Gleichungen

1) eine Diagonalmatrix ist, die

2) auf der Hauptdiagonalen G einander gleiche Untermatrizen V_{11} enthält, die ihrerseits wiederum

---

1) Nur in diesem Falle läßt sich die schwierige Schätzformel der dreistufigen Methode auf die Schätzformel der gewöhnlichen Methode der kleinsten Quadrate zurückführen, denn nur dann kürzt sich σ_{11} gegen $\frac{1}{\sigma_{11}}$ und die Streuungs-Kovarianz-Matrix der latenten Variablen hat keinen Einfluß mehr auf die Schätzung.

a) Diagonalmatrizen sind, welche

b) auf ihrer Hauptdiagonalen nur Element enthalten, die einem bestimmter Skalar σ_{11} gleich sind.

22. Ein reduzierte-Form-Gleichungssystem, das verzögerte endogene Variablen enthält.

Verzögerte endogene Variablen sind für die empirische Wirtschaftsforschung von entscheidender Bedeutung, da wirtschaftliche Anpassungsprozesse an veränderte Umweltsbedingungen oft eine größere zeitliche Dimension aufweisen. Wenn diese Anpassungsprozesse eine längere zeitliche Ausdehnung haben als ihre statistische Erfassungsperiode (Monat, Vierteljahr, Halbjahr oder Jahr), so wird eine in einer Periode beginnende Anpassung auch in den nachfolgenden Perioden die Entwicklung der endogenen Variablen affizieren. Gewissermaßen ist die Anpassung der endogenen Variablen an die veränderten ökonomischen Umweltsbedingungen in der Anfangsperiode zu gering oder zu stark[1], so daß die Korrektur der übermäßig großen oder kleinen Anpassungen die endogenen Variablen der Folgeperioden verändert. Anders gewendet bedeutet dies, daß die Veränderungen bzw. das Niveau der endogenen Variablen von den Veränderungen bzw. dem Niveau der endogenen Variablen früherer Perioden abhängig sein können.

Die Berücksichtigung verzögerter endogener Variablen führt jedoch vom ökonometrischen Standpunkt aus zu einem Dilemma: Treten in den Schätzgleichungen verzögerte endogene Variablen auf, so werden einige wichtige Voraussetzungen verletzt und die Schätzeigenschaften verschlechtern sich. Werden die verzögerten endogenen Variablen jedoch nicht in der Schätzgleichung spezifiziert, so werden sie unter Umständen als Störvariablen durch ihren systematischen Einfluß zu einer

1) Gründe hierfür sind Gewohnheiten, institutionelle Starrheiten und sonstige Faktoren, die die Erwartungen der Wirtschaftssubjekte affizieren und deren Verhalten beeinflussen.

Autokorrelation der latenten Variablen führen, die ebenfalls die Schätzeigenschaften verschlechtert.

Für die gewöhnliche Methode der kleinsten Quadrate bedeutet das Auftreten verzögerter endogener Variablen, daß die vorherbestimmten Variablen nicht mehr stochastisch unabhängig sind von den latenten Variablen, so daß die Schätzungen verzerrt[1] sind. Johnston[2] bemerkt, daß die Schätzungen zumindest noch konsistent sein können, wenn die latenten Variablen normal verteilt sind.

Des weiteren zeigt Johnston[3], daß der Regressionskoeffizient einer verzögerten endogenen Variablen im allgemeinen unterschätzt wird, wenn er ein positives Vorzeichen hat. Dagegen wird dieser Regressionskoeffizient überschätzt[4], wenn zudem noch Autokorrelation der latenten Variablen vorliegt.

Diese Ausführungen gelten gleichfalls für die indirekte Methode der kleinsten Quadrate bei einem reduzierte-Form-Gleichungssystem, da in diesem Falle die Schätzungen dieser Methode mit den Schätzungen nach der gewöhnlichen Methode der kleinsten Quadrate übereinstimmen[5].

Da auch die zweistufige Methode der kleinsten Quadrate bei einem reduzierte-Form-Gleichungssystem mit der gewöhnli-

---

1) Vgl. auch Johnston: a.a.O., S. 212.

2) Johnston: a.a.O., S. 212.

3) Johnston: a.a.O., S. 214-216.

4) Johnston: a.a.O., S. 216. Auch Johnston kann nur die Richtung, nicht aber das Ausmaß der Verzerrung bestimmen.

5) Vgl. auch im Kapitel 21. die Bemerkungen zur indirekten Methode der kleinsten Quadrate.

chen Methode der kleinsten Quadrate übereinstimmt[1], gelten die Ausführungen zur gewöhnlichen Methode der kleinsten Quadrate auch für die zweistufige Methode. Hinzuzufügen ist, daß analog zum Vorgehen der dreistufigen Methode[2] die verzögerten endogenen Variablen aufgrund aller vorherbestimmten Variablen des Gleichungssystems neugeschätzt werden könnten, um die Schätzeigenschaften dieser Methode zu verbessern.

Die verallgemeinerte Methode der kleinsten Quadrate liefert Schätzergebnisse, die sich von denen der gewöhnlichen Methode der kleinsten Quadrate unterscheiden, wenn Autokorrelation[3] der latenten Variablen vorliegt. Sollte etwa durch Nichtberücksichtigung verzögerter endogener Variablen (Spezifikationsfehler) Autokorrelation der latenten Variablen vorliegen, so kann die verallgemeinerte Methode der kleinsten Quadrate zu besseren Schätzergebnissen führen als die gewöhnliche Methode der kleinsten Quadrate, wenn dieser Spezifikationsfehler durch eine bessere Spezifikation der Streuungs-Kovarianz-Matrix der latenten Variablen (hier wäre eine Skalarmatrix eine Fehlspezifikation) ausgeglichen wird.

Die dreistufige Methode der kleinsten Quadrate wird ebenfalls durch das Auftreten verzögerter endogener Variablen in ihren Schätzqualitäten beeinträchtigt. Verzögerte endogene Variablen werden wie die unverzögerten endogenen und die exogenen Variablen einer Gleichung zunächst aufgrund aller vorherbestimmten Variablen des Gleichungssystems

---

1) Vgl. im Kapitel 21. die Bemerkungen zur zweistufigen Methode der kleinsten Quadrate.

2) Vgl. die Betrachtungen im Anschluß an Gleichung (6.12).

3) Vgl. die Betrachtungen im Anschluß an Gleichung (4.24).

neugeschätzt, ehe in der dritten Stufe alle Gleichungen zugleich geschätzt werden. Die Problematik jener Neuschätzung besteht jedoch darin, daß in der Gesamtheit der vorherbestimmten Variablen des Gleichungssystems auch verzögerte endogene Variablen enthalten sind, welche nicht stochastisch unabhängig sind von den latenten Variablen. Das eigentliche Ziel der Neuschätzung, nämlich die Beseitigung der stochastischen Abhängigkeit der neuzuschätzenden Variablen, wird solange nicht erreicht, als die zur Schätzung herangezogenen (Hilfs-)Variablen selbst stochastisch abhängig sind von den latenten Variablen. Hinzu kommt, daß die dreistufige Methode der kleinsten Quadrate das Fehlen von Autokorrelation voraussetzt, was die Anwendbarkeit dieser Methode wesentlich beeinträchtigt. Für den Fall verzögerter endogener Variablen bedeutet diese Methode daher keinen eindeutigen Fortschritt gegenüber der gewöhnlichen Methode der kleinsten Quadrate.

Für die Methode der größten Dichte bei vollständiger Information haben Koopmans und Hood[1] in einem interessanten Beweis gezeigt, daß es für ein reduzierte-Form-Gleichungssystem belanglos ist, ob die vorherbestimmten Variablen verzögerte endogene Variablen enthalten oder nur exogene Variablen. Allerdings ist auch bei diesem Beweis darauf zu achten, daß die latenten Variablen nicht autokorreliert sein dürfen. Wir betrachten die logarithmische Likelihoodfunktion (8.17) für den Fall der reduzierten Form, in welchem sie folgendes Aussehen hat:

$$(22.1) \quad L = -\frac{G}{2}\ln(2\pi) - \frac{1}{2}\ln(\det \Sigma) - \frac{1}{2}\,\mathrm{Sp}\{\Sigma^{-1}[I \quad -\Pi]\, M_{xx}[I \quad -\Pi]'\}$$

1) Koopmans u. Hood: a.a.O., S. 151-155, s. bes. S. 154 sowie S. 124. Vgl. auch Koopmans, Rubin u. Leipnik: a.a.O., S. 115-117.

Das Argument von Koopmans und Hood[1] lautet nun, daß es für die Form von (22.1) ohne Einfluß ist, ob in der Momentenmatrix M_{xx} verzögerte endogene Variablen enthalten sind oder nicht. Die Tatsache, daß die endogenen Variablen autokorreliert sind, wirkt sich nach Koopmans und Hood weder in einer Autokorrelation[2] der latenten Variablen noch in einer stochastischen Verbundenheit[3] zwischen vorherbestimmten und latenten Variablen aus, so daß in der Tat die Likelihoodfunktion nicht affiziert wird durch das Auftreten verzögerter endogener Variablen.

Das Argument von Koopmans und Hood[4] gilt entsprechend für ein Untersystem oder eine einzelne Gleichung des reduzierte-Form-Gleichungssystems.

---

1) Koopmans u. Hood: a.a.O., S. 153-154.
2) Koopmans u. Hood: a.a.O., S. 152
3) Koopmans u. Hood: a.a.O., S. 124 u. 122.
4) Koopmans u. Hood: a.a.O., S. 154-155.

23. Eine einzige Gleichung mit mehr als einer unverzögerten endogenen Variablen.

Wir betrachten vorerst nur den Fall, daß keine verzögerten Variablen auftreten. Treten mehr als eine unverzögerte endogene Variable in einer einzigen, zu schätzenden Gleichung auf, so wird der Vergleich der Schätzergebnisse verschiedener Methoden schwieriger. Die Gleichung, die bereits nach einer der endogenen Variablen, etwa y_1, durch Normierung des zugehörigen Parameters $\beta_{11} = 1$ aufgelöst ist, hat folgendes Aussehen

$$(23.1)\quad y_{1t} = -\beta_{12}y_{2t} - \cdots - \beta_{1G^\Delta}y_{G^\Delta t} - \gamma_{11}x_{1t} - \cdots - \gamma_{1K^*} \cdot x_{K^*t} + u_{1t} \qquad (t=1,\ldots,T).$$

Durch Einsetzen des $(G^\Delta + K^*)$-tupels von Beobachtungswerten für insgesamt T Beobachtungsperioden erhalten wir für (23.1) in Matrizenschreibweise

$$(23.2)\quad y_1 = -Y_2\,\beta_{12}^{\prime} - X_*\,\gamma_{1*}^{\prime} + u_1$$

wobei

$$(23.3)\quad y_1 = \begin{bmatrix} y_{11} \\ \cdot \\ \cdot \\ \cdot \\ y_{1T} \end{bmatrix}, \quad Y_2 = \begin{bmatrix} y_{21} & \cdots & y_{G^\Delta 1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ y_{2T} & \cdots & y_{G^\Delta T} \end{bmatrix},$$

$$\beta_{12}^{\prime} = \begin{bmatrix} \beta_{12} \\ \cdot \\ \cdot \\ \cdot \\ \beta_{1G^\Delta} \end{bmatrix}, \quad X_* = \begin{bmatrix} x_{11} & \cdots & x_{K^*1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_{1T} & \cdots & x_{K^*T} \end{bmatrix},$$

$$Y_{1*}' = \begin{bmatrix} y_{11} \\ \cdot \\ \cdot \\ \cdot \\ y_{1K*} \end{bmatrix} \text{ und } u_1 = \begin{bmatrix} u_{11} \\ \cdot \\ \cdot \\ \cdot \\ u_{1T} \end{bmatrix} .$$

Wir bilden nun aus (23.2)

$$(23.4) \qquad y_1 = - Z_1 \delta_1' + u_1,$$

wobei

$$(23.5) \qquad Z_1 = [Y_2 \quad X_*] \text{ und } \delta_1' = \begin{bmatrix} \beta_{12}' \\ \gamma_{1*}' \end{bmatrix} .$$

Dann lautet die Schätzung von (23.4) mit der gewöhnlichen Methode der kleinsten Quadrate

$$(23.6) \qquad \hat{\delta}_1 = - (Z_1' Z_1)^{-1} Z_1' y_1$$

mit der Verzerrung[1)]

$$(23.7) \qquad E(\hat{\delta}_1') - \delta_1' = E[-(Z_1' Z_1)^{-1} Z_1' u_1]$$

Diese Verzerrung verschwindet nicht - auch asymptotisch nicht, d.h. mit zunehmender Zahl der Beobachtungsperioden[2)]

---

1) Die Verzerrung (23.7) erhält man in gewohnter Weise durch Einsetzen von (23.4) in (23.6) und Bildung der Differenz der mathematischen Erwartung hiervon mit dem "wahren" Parametervektor δ_1.

2) Somit können die Schätzungen mit der gewöhnlichen Methode der kleinsten Quadrate nicht konsistent sein. Vgl. auch Theil: Economic Forecast and Policy. a.a.O., S. 232.

- da $E(Z_1' u_1) \neq 0$, weil die Untermatrix von Z_1, nämlich Y_2, ebenso von dem Vektor der latenten Variablen, u_1, abhängt wie y_1, der Vektor der ausgewählten endogenen Variablen.

Da andererseits die Erwartungstreue eine notwendige Bedingung für die Effizienz einer Schätzung ist, können wir folgern, daß die gewöhnliche Methode der kleinsten Quadrate nicht zu effizienten Schätzungen führen kann, wenn in einer einzigen Gleichung mehrere unverzögerte endogene Variablen auftreten.

Da die Verzerrung auch asymptotisch, d.h. mit zunehmendem Stichprobenumfang, nicht verschwindet, ist die gewöhnliche Methode der kleinsten Quadrate in diesem Fall noch nicht einmal konsistent. Damit erfüllt die Methode der kleinsten Quadrate hier nicht einmal die "Minimalforderung, die man an die Güte einer Schätzung stellt".[1)]

Die indirekte Methode der kleinsten Quadrate als reduzierte-Form-Schätzmethode erfordert für eine einzige Gleichung mit mehr als einer unverzögerten endogenen Variablen, daß wir die reduzierte-Form-Gleichung, die nach y_1 aufgelöst ist, berechnen können. Hierzu ist es allerdings notwendig, daß wir auch die anderen Gleichungen des Gleichungssystems kennen, obschon wir an deren Schätzung eventuell nicht interessiert sind. Wir benötigen ein vollständiges Modell, bestehend aus Strukturgleichungen, um hieraus das zugehörige Modell, bestehend aus reduzierte-Form-Gleichungen, berechnen zu können. Um die reduzierte-Form-Gleichung

1) Menges: a.a.O., S. 89. Zu diesem Ergebnis kommen auch Koopmans und Hood (a.a.O., S. 134).

schätzen zu können, ist es notwendig, daß sie identifizierbar[1] ist.

Die indirekte Methode der kleinsten Quadrate ist nur bei exakter Identifikation anwendbar und führt dann und nur dann zu identischen Ergebnissen wie die Methode der größten Dichte bei beschränkter Information. Bei exakter Identifikation ist mit Hilfe einer Normierung[2] eine Rückrechnung der reduzierte-Form-Schätzparameter in Schätzparameter der Strukturgleichung (23.1) noch möglich. Dagegen müßte bei Überidentifikation ein weiterer Freiheitsgrad durch eine weitere Festsetzung beseitigt werden. Eine solche weitere Festsetzung ist jedoch durch die Strukturgleichung (23.1) nicht a priori fixiert, so daß ihre nachträgliche Einführung nur noch zu willkürlichen Ergebnissen führen könnte. Bleibt ohne eine solche willkürliche Festsetzung der Freiheitsgrad bestehen, so lassen sich allenfalls Verhältnisse der Parameterschätzungen $\hat{\beta}_{1i}$, wobei $(i=2,\ldots,G^{\Delta})$, zueinander bilden, den absoluten Wert $\hat{\beta}_{1i}$ für $(i=2,\ldots, G^{\Delta})$ jeder Schätzung erhalten wir jedoch nicht.

Die verallgemeinerte Methode der kleinsten Quadrate zeichnet sich durch die explizite Berücksichtigung der Streuungs-Kovarianz-Matrix der latenten Variablen bei der Schätzung aus. Dagegen führt diese Methode wie die gewöhnliche Methode der kleinsten Quadrate zu verzerrten Schätzungen, wenn in einer Gleichung mehrere unverzögerte Variablen enthalten sind. Diese Verzerrung ergibt sich daraus, daß eine endogene Variable in Abhängigkeit von anderen endogenen Variablen geschätzt wird, obwohl alle endogenen Variablen zugleich

---

1) Vgl. zur Identifikation die Darstellung der Methode der größten Dichte bei beschränkter Information nach Anderson und Rubin (Kapitel 9).

2) Gemeint ist die übliche Normierung $\beta_{11}=1$, da (23.1) nach der endogenen Variablen y_{1t} aufgelöst ist.

von derselben latenten Variablen stochastisch abhängig sind. Dieses Vorgehen entspricht einer Fehlspezifikation der zur Schätzung herangezogenen endogenen Variablen als "stochastisch unabhängige vorherbestimmte Variablen". Da diese Fehlspezifikation sich nicht direkt in der Streuungs-Kovarianz-Matrix der latenten Variablen auswirkt, vermag auch eine noch so gute Spezifikation dieser Matrix nicht die Konsequenzen der Fehlspezifikation der "erklärenden" endogenen Variablen zu beseitigen, so daß also auch die verallgemeinerte Methode der kleinsten Quadrate nicht einmal zu konsistenten Schätzungen führt, da die Verzerrung auch asymptotisch nicht verschwindet.

Die zweistufige Methode der kleinsten Quadrate liefert für eine einzige Gleichung mit mehreren unverzögerten endogenen Variablen konsistente[1] Schätzungen[2]

$$(23.8)\qquad \hat{\delta}_1' = -\begin{bmatrix} Y_2' Y_2 - w'w & Y_2' X_* \\ X_*' Y_2 & X_*' X_* \end{bmatrix}^{-1} \begin{bmatrix} Y_2' - w' \\ X_*' \end{bmatrix} y_1 .$$

Somit ist die zweistufige Methode der kleinsten Quadrate den in diesem Fall bisher betrachteten Methoden vorzuziehen. Dies verwundert nicht, da die zweistufige Methode der kleinsten Quadrate als Ein-Gleichungs-Methode gerade für diesen Fall des Auftretens mehrerer unverzögerter endogener Variablen in einer Gleichung von Henri Theil konzipiert worden ist. Hiermit erreichte Theil gewissermaßen einen Anschluß der Kleinst-Quadrate-Methode an kompliziertere Verfahren, die - wie z.B. die Methode der größten

1) Vgl. Theil (1961): a.a.O., S. 232.
2) Vgl. Gleichung (5.34).

Dichte bei beschränkter Information - bereits in diesem Fall zu konsistenten[1] Schätzungen führten.

Die dreistufige Methode der kleinsten Quadrate liefert für eine einzige Gleichung mit mehreren unverzögerten endogenen Variablen konsistente Schätzungen, da die dreistufige Methode als Fortentwicklung der zweistufigen Methode der kleinsten Quadrate deren Vorzüge beibehält. Die Ergebnisse der zweistufigen und der dreistufigen Methode sind identisch, wenn die Kovarianz-Matrizen v_{ij} für $i \neq j$ $(i, j = 1, \ldots, G)$ Nullmatrizen sind[2].

Die Methode der größten Dichte bei vollständiger Information, bei der die latenten Variablen als normal verteilt vorausgesetzt werden, erzielt auch bei einer Gleichung mit mehreren unverzögerten endogenen Variablen konsistente[3] Schätzungen. Dies ist die Folge der gleichmäßigen Behandlung aller unverzögerten endogenen Variablen dieser Gleichung bei der Maximierung der logarithmierten Dichtefunktion der endogenen Variablen bezüglich aller zu schätzenden Parameter der Gleichung. Im Gegensatz hierzu würde eine "naive" Anwendung der gewöhnlichen Methode der kleinsten Quadrate eine der endogenen Variablen als durch die vorherbestimmten und auch durch die übrigen unverzögerten endogenen Variablen dieser Gleichung erklärt betrachten. Die Schätzung in direkter Abhängigkeit von den vorherbestimmten und den übrigen unverzögerten endogenen Variablen würde einmal auf der Fehlspezifikation der stochasti-

---

1) Vgl. Theil: Economic Forecasts and Policy: a.a.O., S. 232.

2) Vgl. hierzu S. 57.

3) Vgl. Menges (1961): a.a.O., S. 104 sowie das 8. Kapitel dieser Arbeit.

schen Eigenschaften der übrigen unverzögerten endogenen Variablen beruhen. Zum anderen würde der Erklärungsgehalt der übrigen Gleichungen vernachlässigt, in denen weitere vorherbestimmte Variablen bei Vorliegen von Identifikation gerade jene übrigen unverzögerten endogenen Variablen beeinflussen. Gerade weil die Methode der größten Dichte bei vollständiger Information diese Fehlerquellen vermeidet und den Gesamtaufbau des Modells mit allen seinen Interdependenzen berücksichtigt, führt sie hier im Gegensatz zur gewöhnlichen Methode der kleinsten Quadrate zu konsistenten Schätzungen.

Ebenfalls unter der Annahme normal verteilter latenter Variablen sind die Schätzungen der Methode der größten Dichte bei beschränkter Information konsistent[1] für eine Gleichung mit mehreren unverzögerten endogenen Variablen. Allerdings ist die Methode der größten Dichte bei vollständiger Information asymptotisch effizient[2] im Vergleich zur Methode der größten Dichte bei beschränkter Information, da letztere Methode nicht die a-priori-Restriktionen der übrigen Gleichungen eines Systems berücksichtigt, sondern nur die Beobachtungswerte der aus der Schätzgleichung ausgeschlossenen endogenen und vorherbestimmten Variablen des Systems verwendet, gleichgültig, in welchen der übrigen Gleichungen diese Variablen auftreten oder a priori ausgeschlossen sind.

---

1) Vgl. Chernoff, H. u. Rubin, H.: Asymptotic properties of limited-information estimates under generlized conditions. In: Studies in Econometric Method. Hrsg.: W.C. Hood und Tjalling C. Koopmans. New York-London 1953, S. 200-212, s. bes. S. 201-202.

2) Vgl. Menges: a.a.O., S. 106 u. 104.
Vgl. auch Anderson, T.W.: Estimation of the parameters of a single equation by the limited-information maximum-likelihood method. In: Statistical Inference in Dynamic Economic Models. Hrsg.: Tjalling C. Koopmans. New York-London 1950, S. 311-322, s. bes. S. 312; vgl. auch S. 316-317 zur Konsistenz.

Treten verzögerte endogene Variablen auf, so darf wie bei der reduzierten Form mit Verzerrungen gerechnet werden, die von der Art der stochastischen Abhängigkeit dieser Variablen abhängig sind. Das Nebeneinander von mehreren unverzögerten endogenen Variablen, stochastisch abhängigen Variablen und Autokorrelation der latenten Variablen ist jedoch derart kompliziert, daß es noch umfangreicher ökonometrischer Forschung, und zwar theoretischer wie empirischer Natur bedarf, ehe die wichtigsten Zusammenhänge entdeckt werden können. Um nur auf die Teilaspekte dieses Komplexes aufmerksam zu machen, sei auf die Stellungnahmen von Koopmans und Hood[1] einerseits und Johnston[2] andererseits verwiesen.

1) Koopmans u. Hood: a.a.O., S. 133-134.
2) Johnston: a.a.O., S. 216.

24. Ein interdependentes Gleichungssystem.

Wir betrachten ein Gleichungssystem, in welchem keine verzögerten endogenen Variablen enthalten sind.[1] Dieses Gleichungssystem soll aus interdependenten[2] Strukturgleichungen bestehen, d.h. aus Strukturgleichungen, in denen die endogenen Variablen durch ihr gleichzeitiges Auftreten in verschiedenen Strukturgleichungen gegenseitig abhängig sind. Der Spezialfall des rekursiven Gleichungssystems, bei dem sich wegen der Triangularität der Parametermatrix der endogenen Variablen nach Wold[3] eine einseitige kausale Abhängigkeit der endogenen Variablen herleiten läßt, ist formal[3] als Unterfall des interdependenten Gleichungssystems zu betrachten und wird somit implizit mitbehandelt.

Unser interdependentes Gleichungssystem habe folgendes Aussehen

$$(24.1) \qquad B\, y_t' + \Gamma\, x_t' = u_t' \qquad (t=1,\ldots,T)^{4},$$

wobei

$$(24.2) \quad B = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1G} \\ \vdots & & \vdots \\ \beta_{G1} & \cdots & \beta_{GG} \end{bmatrix}, \; y_t' = \begin{bmatrix} y_{1t} \\ \vdots \\ y_{Gt} \end{bmatrix},$$

$$\Gamma = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1K} \\ \vdots & & \vdots \\ \gamma_{G1} & \cdots & \gamma_{GK} \end{bmatrix}, \; x_t' = \begin{bmatrix} x_{1t} \\ \vdots \\ x_{Kt} \end{bmatrix} \text{ und } u_t' = \begin{bmatrix} u_{1t} \\ \vdots \\ u_{Gt} \end{bmatrix}.$$

1) Die Problematik verzögerter endogener Variablen wird hier vernachlässigt, da die am Ende des vorangehenden Kapitels geäußerten Bemerkungen auch für dieses Kapitel Geltung haben.
2) Vgl. Menges (1961): a.a.O., S. 51-52.
3) Wold, H. (in Zusammenarbeit mit Juréen, L.): a.a.O., S. 50.
4) Im folgenden wird (t=1,...,T) nicht mitgeführt; gleichwohl soll der mitgeführte Index t die Möglichkeit der Zusammenfassung der Beobachtungswerte zu Matrizen und Vektoren vergegenwärtigen.

Die Strukturgleichungen dieses Systems oder die hieraus zu berechnenden reduzierte-Form-Gleichungen sollen wieder - je nach Eigenschaft des Verfahrens - von alternativen Schätzmethoden geschätzt werden. Beginnen wir mit der gewöhnlichen Methode der kleinsten Quadrate. Die "naive" Anwendung dieser Methode zur Schätzung des Gleichungssystems (24.1) bestünde darin, jede der G Gleichungen nach einer anderen endogenen Variablen aufzulösen, wobei dann die restlichen endogenen Variablen auf der rechten Seite der Gleichung stehen würden. So erhielten wir bei systematischem Vorgehen z.B. statt der ersten Gleichung des Gleichungssystems (24.1), nämlich

$$(24.3)\quad \beta_{11}y_{1t} + \beta_{12}y_{2t} + \cdots + \beta_{1G}y_{Gt} + \gamma_{11}x_{1t} + \cdots$$
$$\cdots + \gamma_{1K}x_{Kt} = u_{1t}$$

die Gleichung

$$(24.4)\quad y_{1t} = -\frac{\beta_{12}}{\beta_{11}}y_{2t} - \cdots - \frac{\beta_{1G}}{\beta_{11}}y_{Gt} - \frac{\gamma_{11}}{\beta_{11}}x_{1t} - \cdots$$
$$\cdots - \frac{\gamma_{1K}}{\beta_{11}}x_{Kt} + \frac{u_{1t}}{\beta_{11}}\,.$$

Wir sehen, daß an die Stelle der bisherigen Matrizen B und Γ die Matrizen B* und Γ* treten, wobei die Hauptdiagonale der Matrix B* aus Einsen besteht, während die übrigen Elemente von B* und die Elemente von Γ* Verhältnisse sind, die wir durch folgende Multiplikationen erhalten

$$(24.5)\quad B^* = \begin{bmatrix} \frac{1}{\beta_{11}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\beta_{22}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{\beta_{GG}} \end{bmatrix} B$$

und

$$(24.6)\quad \Gamma^* = \begin{bmatrix} \frac{1}{\beta_{11}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\beta_{22}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{\beta_{GG}} \end{bmatrix} \Gamma$$

An die Stelle des Spaltenvektors u_t' tritt der Spaltenvektor $u_t^{*\prime}$, den wir erhalten durch

$$(24.7)\quad u_t^{*\prime} = \begin{bmatrix} \frac{1}{\beta_{11}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\beta_{22}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{\beta_{GG}} \end{bmatrix} u_t'$$

Wir erhalten also B* durch Multiplikation jeder Zeile von B mit dem Kehrwert des Hauptdiagonalelements dieser Zeile von B. Die Matrix Γ^* erhalten wir durch Multiplikation der i-ten Zeile von Γ wobei (i=1,...,G) mit dem Kehrwert des Hauptdiagonalelements der i-ten Zeile von B (wobei ebenfalls (i=1,...,G).

Wenn wir die latente Variable der i-ten Gleichung, wobei (i=1,...,G), durch das Hauptdiagonalelement der i-ten Zeile, wobei (i=1,...,G) von B dividieren, erhalten wir analog zu (24.4) nach einer kleinen Umformung auch die nach $y_{2t}, y_{3t}, \ldots, y_{Gt}$ aufgelösten Gleichungen. Werden diese Gleichungen [von der Art der Gleichung (24.4)] nach der gewöhnlichen Methode der kleinsten Quadrate geschätzt, so tritt folgender Schätzfehler auf: Die nicht ausgewählten endogenen Variablen auf der rechten Seite dieser Gleichungen [in Gleichung (24.4) also die Variablen $y_{2t}, \ldots, y_{Gt}$] sind ebenfalls von der latenten Variablen [in Gleichung (24.4) also von der latenten Variablen $\frac{u_{1t}}{\beta_{11}}$] abhängig. Damit ist die fundamentale Voraussetzung der Methode der kleinsten Quadrate verletzt, nämlich, daß die Variablen, mit deren Hilfe eine andere Variable geschätzt werden soll, von der latenten Variablen stochastisch unabhängig sein müsse. Die sich aus der Verletzung dieser Voraussetzung ergebende Verzerrung wollen wir für die Gleichung (24.4) berechnen.

Aus (24.4) bilden wir

$$(24.8) \quad y_{1t} = -\beta_2^* Y_{2t}' - \gamma_1^* x_t + u_{1t}^*,$$

wobei

$$(24.9)\quad B_2^* = \begin{bmatrix} \frac{\beta_{12}}{\beta_{11}} & \frac{\beta_{13}}{\beta_{11}} & \cdots & \frac{\beta_{1G}}{\beta_{11}} \end{bmatrix}, \quad Y_{2t}' = \begin{bmatrix} y_{2t} \\ \vdots \\ y_{Gt} \end{bmatrix},$$

$$\Gamma_1^* = \begin{bmatrix} \frac{\gamma_{11}}{\beta_{11}} & \cdots & \frac{\gamma_{1K}}{\beta_{11}} \end{bmatrix}, \quad u_{1t}^* = \frac{u_{1t}}{\beta_{11}}$$

und x_t' wie in (24.2) definiert ist.

Für (24.8) können wir weiter schreiben

$$(24.10)\quad y_{1t} = -\delta_1^* z_t' + u_{1t}^*,$$

wobei

$$(24.11)\quad \delta_1^* = [B_2^* \quad \Gamma_1^*] \text{ und } z_t' = \begin{bmatrix} Y_{2t}' \\ x_t' \end{bmatrix}.$$

Dann ist die Kleinst-Quadrat-Schätzung von δ_1^* [1)]

$$(24.12)\quad \hat{\delta}_1^* = -\, y_{1t}'\, z_t\, (z_t'\, z_t)^{-1}$$

mit der Verzerrung[2)]

$$(24.13)\quad E(\hat{\delta}_1^*) - \delta_1^* = E[-\, u_{1t}^{*\prime}\, z_t (z_t'\, z_t)^{-1}].$$

---

1) Hierbei ist für die erste Strukturgleichung y_{1t}' ein Zeilenvektor der Ordnung $(1 \times T)$ und z_t die Matrix der Beobachtungen der Ordnung $[T \times (G - 1 + K)]$.

2) Diese berechnen wir in gewohnter Weise mit Hilfe von (24.10) und (24.12).

Diese Verzerrung verschwindet auch mit zunehmender Zahl der Beobachtungsperioden nicht, so daß die gewöhnliche Methode der kleinsten Quadrate hier nicht einmal zu konsistenten Schätzungen führt.

Daß dieses Ergebnis für alle Gleichungen gilt, sieht man, wenn man aus (24.10) das transformierte[1] (d.h. nach jeweils einer der endogenen Variablen aufgelöste) Strukturgleichungssystem bildet

$$(24.14) \qquad y_t' = - D^* Z_t' + u_t^{*\prime} ,$$

wobei y_t' wie in (24.2) ist und

$$(24.15) \qquad D^* = [B^{*(2)} \quad \Gamma^*], \; Z_t' = \begin{bmatrix} y_t' \\ x_t' \end{bmatrix}$$

und $u_t^{*\prime}$ durch (24.7) definiert ist.

Hierbei ist neu

$$(24.16) \qquad B^{*(2)} = B^* - I_G ,$$

das aus B* entsteht, wenn man die Einheitsmatrix der Ordnung G subtrahiert.

Statt der Verzerrung von (24.13) erhalten wir dann die Verzerrung

$$(24.17) \qquad E(\hat{D}^*) - D^* = E[- u_t^{*\prime} Z_t (Z_t' Z_t)^{-1}],$$

---

1) Nicht zu verwechseln mit dem reduzierte-Form-Gleichungssystem, bei dem auf der rechten Seite der Gleichungen keine endogenen Variablen mehr auftreten.

welche ebenfalls asymptotisch nicht verschwindet und zeigt, daß die Methode der kleinsten Quadrate bei der Schätzung eines Systems von Strukturgleichungen nicht zu konsistenten Schätzungen führt.

Die indirekte Methode der kleinsten Quadrate führt zu konsistenten Schätzungen, wenn aus dem Gleichungssystem (24.1) ein reduzierte-Form-Gleichungssystem gebildet werden kann, welches exakt identifizierbar[1] ist.

Die reduzierte Form läßt sich aus (24.1) bilden, wenn die Kehrmatrix von B gebildet werden kann. Die Kehrmatrix von B läßt sich jedoch nur bilden, wenn B regulär ist, d.h. wenn die Determinante der Matrix B von Null verschieden ist.

Die Parameter der durch Multiplikation von (24.1) von links mit der Kehrmatrix von B errechneten reduzierten Form lassen sich jedoch nur schätzen, wenn alle Gleichungen der reduzierten Form identifizierbar[1] (also exakt- oder überidentifizierbar) sind. Dagegen ist die Berechnung der Parameterschätzungen der Strukturgleichungen von System (24.1) dann und nur dann aus den Parameterschätzungen der reduzierte-Form-Gleichungen möglich, wenn jede der Strukturgleichungen des Strukturgleichungssystems exakt identifizierbar[2] ist. Da dieser Fall in der ökonometrischen Praxis selten vorkommt, ist die indirekte Methode der kleinsten Quadrate im allgemeinen zur Schätzung eines Gleichungssystems von Strukturgleichungen ungeeignet.

Die verallgemeinerte Methode der kleinsten Quadrate führt bei einem Gleichungssystem von Strukturgleichungen, wobei

---

1) Vgl. auch S. 58.

2) Vgl. die Betrachtungen zur Identifikation im 9. Kapitel.

in den Gleichungen jeweils mehrere unverzögerte endogene Variablen auftreten, nicht zu konsistenten Schätzungen. Zwar kann die stochastische Verbundenheit der latenten Variablen verschiedener Gleichungen durch die Spezifikation der Varianz-Kovarianz-Matrix der latenten Variablen des Gleichungssystems berücksichtigt werden. Dennoch wird der gleiche Fehler wie bei der "naiven" Anwendung der gewöhnlichen Methode der kleinsten Quadrate auf ein solches Gleichungssystem begangen, nämlich: Eine endogene Variable wird in Abhängigkeit von vorherbestimmten Variablen und von anderen, ebenfalls von der latenten Variablen der Gleichung beeinflußten endogenen Variablen geschätzt. Die fundamentale Voraussetzung der Methode der kleinsten Quadrate, daß die beobachteten Variablen $Z_t^!$ auf der rechten Seite des Systems (24.14) von den latenten Variablen $u_t^{*'}$ des Systems (24.14) stochastisch unabhängig sein sollen, ist jedoch verletzt, weil $Z_t^!$ auch die endogenen Variablen $y_t^!$ enthält. Die resultierende Verzerrung der Schätzung verschwindet auch asymptotisch nicht, so daß die Schätzungen nicht konsistent sind.

Die zweistufige Methode der kleinsten Quadrate würde jede einzelne Gleichung des Gleichungssystems (24.14) unabhängig von den anderen Gleichungen des Systems (24.14) schätzen. Es wäre an dieser Stelle auch zu berücksichtigen, daß für die Koeffizientenmatrizen $B^{*(2)}$ und Γ^*, die zusammen die Koeffizientenmatrix D^* bilden, a-priori-(Null-)Restriktionen gelten. Nach der zweistufigen Methode der kleinsten Quadrate werden dann in der ersten Stufe die in einer Gleichung enthaltenen vorherbestimmten und nicht-ausgewählten endogenen Variablen geschätzt aufgrund aller vorherbestimmten Variablen des Systems. Da angenommen wird, daß die vorherbestimmten Variablen des Systems fest vorgegeben und daher von den latenten Variablen stochastisch unabhängig sind, sind auch die Schätzungen aufgrund der vorherbestimm-

ten Variablen stochastisch unabhängig von den latenten Variablen. Dann führt die zweite Stufe der zweistufigen Methode der kleinsten Quadrate zu konsistenten Schätzungen, da die neugeschätzten vorherbestimmten und nicht ausgewählten endogenen Variablen jeder Gleichung von der latenten Variablen der Gleichung unabhängig sind.

Die dreistufige Methode der kleinsten Quadrate führt zu effizienten Schätzungen im Vergleich zur zweistufigen Methode der kleinsten Quadrate, weil sie die Streuungs-Kovarianz-Matrix der latenten Variablen aller Gleichungen berücksichtigt.

Die dreistufige Methode der kleinsten Quadrate setzt insbesondere voraus, daß die latenten Variablen der einzelnen Gleichungen für verschiedene Perioden nicht korreliert sind und eine zeitlich konstante Streuung besitzen. Dagegen ist nicht erforderlich, daß die Kovarianzen der latenten Variablen verschiedener Strukturgleichungen zu demselben Zeitpunkt gleich Null sind, sie müssen nur zeitlich konstant sein. Des weiteren müssen zu einem bestimmten Zeitpunkt die Kovarianzen der latenten Variablen verschiedener Strukturgleichungen nicht für alle Gleichungspaare identisch sein[1]. Dann ergibt sich folgendes Aussehen der symmetrischen Streuungs-Kovarianz-Matrix aller latenten Variablen des Systems von Strukturgleichungen (24.1)

$$(24.18)\quad V = E(u\,u') = \begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1G} \\ V_{21} & V_{22} & \cdots & V_{2G} \\ \vdots & \vdots & & \vdots \\ V_{G1} & V_{G2} & \cdots & V_{GG} \end{bmatrix},$$

1) Auch die Varianzen der latenten Variablen verschiedener Strukturgleichungen müssen nicht identisch sein.

wobei

$$(24.19)\quad V_{ij} = E \begin{bmatrix} (u_{i1}\ u_{j1})(u_{i1}\ u_{j2}) & \dots & (u_{i1}\ u_{jT}) \\ (u_{i2}\ u_{j1})(u_{i2}\ u_{j2}) & \dots & (u_{i2}\ u_{jT}) \\ \vdots & \vdots & \vdots \\ (u_{iT}\ u_{j1})(u_{iT}\ u_{j2}) & \dots & (u_{iT}\ u_{jT}) \end{bmatrix}$$

$$= \sigma_{ij}\ I_T,$$

da die latenten Variablen verschiedener Strukturgleichungen zu verschiedenen Zeitpunkten (d.h. für t=1,...,T) bzw. t'=1,...,T und t ≠ t' nicht korreliert sein dürfen.

Da die zweistufige Methode die Streuungs-Kovarianz-Matrix aller latenten Variablen V gewissermaßen nur für den Spezialfall

$$V_{ij} = 0 \qquad \text{für } i \neq j,$$
$$V_{ij} = V_{11} \qquad \text{für } i = j$$

und natürlich $V_{ij} = \sigma_{ij}\ I_T$ berücksichtigt, wird die dreistufige Methode wegen ihres allgemeineren Charakters in der Regel zu effizienteren Schätzungen führen.

Die Methode der größten Dichte bei vollständier Information führt auch bei einem Gleichungssystem von Strukturgleichungen zu konsistenten[1] Schätzungen, wenn die latenten Variablen normalverteilt sind und ihre Varianz-Kovarianz-

1) Vgl. Menges (1961): a.a.O., S. 104.

Matrix regulär ist. Sind die latenten Variablen nicht normal verteilt, so können die Quasi-Maximum-Likelihood-Schätzungen gleichwohl unter bestimmten restriktiven Voraussetzungen[1] zu konsistenten Schätzungen führen.

Die Methode der größten Dichte bei vollständiger Information führt im Gegensatz zur "naiven" Anwendung der gewöhnlichen Methode der kleinsten Quadrate deshalb bei einem System von Strukturgleichungen zu konsistenten Schätzungen, weil die verschiedenen stochastischen Eigenschaften der verschiedenen Variablengruppen berücksichtigt werden. Insbesondere betrachtet die Methode der größten Dichte die nicht ausgewählten endogenen Variablen der Gleichungen nicht wie als fest vorgegebene oder stochastisch unabhängige Variablen, sondern berücksichtigt deren Beziehungen zu den latenten Variablen in der Jacobischen Determinante.

Außerdem sind die Schätzungen nach der Methode der größten Dichte bei vollständiger Information zumindest asymptotisch effizient[2], weil die Streuungs-Kovarianz-Matrix der latenten Variablen explizit berücksichtigt und geschätzt wird.

Die Methode der größten Dichte bei beschränkter Information führt auch für ein Gleichungssystem von Strukturgleichungen zu konsistenten[3] Schätzungen. Gleichwohl werden im allgemeinen die Schätzungen der Methode der größten Dichte bei beschränkter Information, die für jede Gleichung oder jedes Untersystem separat durchgeführt werden, eine größere Streuung[4] aufweisen als die Schätzungen der Methode der

1) Vgl. Koopmans, Rubin u. Leipnik: a.a.O., S. 136-155, s. bes. auf S. 139 das Theorem 3.3.6.

2) Vgl. Menges (1961): a.a.O., S. 104.

3) Vgl. Chernoff u. Rubin: a.a.O., S. 201-202.

4) Vgl. Menges (1961): a.a.O., S. 106 u. 104.

größten Dichte bei vollständiger Information. Dies resultiert daraus, daß insgesamt, nach der Gleichung-für-Gleichung-Schätzung, zwar alle Daten berücksichtigt werden, jedoch die wesentlichen Zusammenhänge zwischen den einzelnen Gleichungen, die sich jeweils im Rest der Streuungs-Kovarianz-Matrix der latenten Variablen zeigen, durch Lineartransformationen oder durch das Rechenverfahren selbst ausgeschaltet werden. Da also die Methode der größten Dichte bei vollständiger Information die bei Strukturgleichungen vorhandenen Interdependenzen der Gleichungen und ihrer Variablen berücksichtigt, sollte sie bei der Schätzung aller Gleichungen des Systems der gleichungsweisen Schätzung mit der Methode der größten Dichte bei beschränkter Information vorgezogen werden.[1)]

1) Wenn man von dem höheren Rechenaufwand absieht.

DRITTER TEIL

Synthese

Es war Ziel dieser Arbeit, die verschiedenen Schätzverfahren darzustellen, um eine Diskussionsgrundlage zu schaffen, die keine grundlegenden Irrtümer über Voraussetzungen und Aufbau der Verfahren gestatten sollte. Die wichtigsten dieser Verfahren wurden in der dann folgenden Kritik anhand von vier Hauptanwendungsfällen einander gegenübergestellt. Hierbei haben sich mit zunehmender Komplexität des Anwendungsfalles gravierende Unterschiede der Ergebnisse der einzelnen Verfahren herausgestellt. Die wichtigsten Erkenntnisse des kritischen Teiles sollen hier nochmals in Kurzform zusammengestelt werden.

25. Ergebnisse des kritischen Teils.

25.a. Eine einzige Gleichung mit einer endogenen Variablen

Unter den Koopmans-Hood-Bedingungskonstellationen[1] erzielen folgende Schätzverfahren identische Schätzergebnisse, solange sie nicht weitere Informationen (etwa aus anderen Gleichungen eines größeren Systems) verwenden:

(1) Die gewöhnliche Methode der kleinsten Quadrate,
(2) die indirekte Methode der kleinsten Quadrate,
(3) die zweistufige Methode der kleinsten Quadrate und
(4) die dreistufige Methode der kleinsten Quadrate.

Dagegen ergibt sich eine Identität dieser Schätzergebnisse mit den Schätzergebnissen der Größte-Dichte-Methoden (Methode der größten Dichte bei vollständiger Information und Methode der größten Dichte bei beschränkter Information) nur dann, wenn die Bedingungskonstellationen vorliegen, die mit der Normalverteilungsvoraussetzung der latenten Variablen kombiniert sind. Sind dagegen die latenten Variablen nicht normalverteilt, so kann nicht mehr mit Sicherheit

1) Vgl. Koopmans u. Hood: a.a.O., S. 131-133.

auf eine Identität der Schätzergebnisse geschlossen werden. Im Zweifel sind Größte-Dichte-Schätzungen den übrigen Schätzungen vorzuziehen, wenn die Dichtefunktion korrekt spezifiziert ist.

In diesen Spezialfällen wird die Methode der größten Dichte bei beschränkter Information, die bei einem aus nur einer Gleichung bestehenden Gleichungssystem mit der Methode der größten Dichte bei vollständiger Information hinsichtlich der Schätzergebnisse identisch ist, dieser letzteren Methode wegen der leichteren Rechenbarkeit vorzuziehen sein. Bei normalverteilten latenten Variablen lassen sich dagegen gewöhnliche Kleinst-Quadrat-Schätzungen - wohlgemerkt, nur bei einer einzigen Gleichung mit einer endogenen Variablen - schneller berechnen als andere Schätzungen identischer Schätzeigenschaften.

Die verallgemeinerte Methode stimmt bezüglich der Schätzergebnisse mit den vier erstgenannten Methoden nur überein, wenn keine Autokorrelation und zugleich keine Heteroskedastizität vorliegen.

25.b. Ein reduzierte-Form-Gleichungssystem ohne verzögerte endogene Variablen

Folgende Methoden erzielen identische Schätzergebnisse:

(1) Die gewöhnliche Methode der kleinsten Quadrate,
(2) die indirekte Methode der kleinsten Quadrate und
(3) die zweistufige Methode der kleinsten Quadrate.

Die Größte-Dichte-Methoden (Methode der größten Dichte bei vollständiger Information und Methode der größten Dichte

bei beschränkter Information) liefern in diesem Falle nur dann die gleichen Ergebnisse wie die vorgenannten Methoden, wenn

(a) die latenten Variablen normalverteilt sind,
(b) die latenten Variablen stochastisch unabhängig voneinander sind und einer gemeinsamen Dichtefunktion gehorchen,
(c) die latenten Variablen mit den exogenen Variablen nicht korreliert sind.

Die verallgemeinerte Methode der kleinsten Quadrate führt zu den gleichen Ergebnissen wie die erstgenannten drei Methoden, wenn die Streuungs-Kovarianz-Matrix der latenten Variablen jeder reduzierte-Form-Gleichung für sich eine Skalarmatrix[1] ist. Anders gewendet heißt dies, daß eine Identität nur zustande kommt bei Homoskedastizität und zugleich verschwindenden Kovarianzen der latenten Variablen dieser Gleichung zu verschiedenen Zeitpunkten.

Die dreistufige Methode der kleinsten Quadrate führt zu den gleichen Schätzergebnissen wie die erstgenannten drei Methoden, wenn das zu betrachtende Gleichungssystem ein reduzierte-Form-Gleichungssystem ist. Es ist jedoch zu betonen, daß diese Identität nur für diesen Sonderfall Gültigkeit besitzt.

---

1) Nach Zurmühl: Matrizen a.a.O., S. 12 ist eine Skalarmatrix eine Matrix, die sich aus dem Produkt der Einheitsmatrix mit einem Skalar ergibt. Hiervon ist das Skalarprodukt zu unterscheiden (Produkt aus Zeilenvektor mit nachfolgendem Spaltenvektor).

Wir sehen, daß selbst bei der reduzierten Form ohne verzögerte endogene Variablen sehr differenzierte Ergebnisse erzielt werden können, wenn einige Voraussetzungen nicht erfüllt sind.

25.c. Ein reduzierte-Form-Gleichungssystem mit verzögerten endogenen Variablen

Die Kleinst-Quadrat-Methoden sind gegen das Auftreten verzögerter endogener Variablen im allgemeinen anfälliger als die Größte-Dichte-Methoden. Am stärksten betroffen durch die hieraus resultierende Verzerrung der Schätzungen sind die gewöhnliche Methode und die für ein reduzierte-Form-Gleichungssystem mit ihr übereinstimmende indirekte Methode der kleinsten Quadrate. Während die verallgemeinerte Methode der kleinsten Quadrate und die dreistufige Methode der kleinsten Quadrate nur partiell die Problematik verzögerter endogener Variablen erfassen können, muß die zweistufige Methode durch folgende Variationen erst an das Problem herangeführt werden: Es wäre vorzuziehen, wenn statt der bei der reduzierten Form nicht vorkommenden restlichen endogenen Variablen der Gleichung die verzögerten endogenen Variablen neugeschätzt würden. Als "Hilfsvariablen" sollten jedoch nicht alle vorherbestimmten Variablen des Gleichungssystems sondern nur alle exogenen Variablen, nicht also die verzögerten endogenen Variablen, des Gleichungssystems herangezogen werden. Durch diese Variation ist die Erzielung konsistenter Schätzungen eher möglich als durch eine unkritische Anwendung der zweistufigen Methode der kleinsten Quadrate. Eine ähnliche Variation der dreistufigen Methode der kleinsten Quadrate wäre m.E. erst sinnvoll, wenn die Variation der zweistufigen Methode sich in empirischen Untersuchungen bewährt haben würde.

25.d. Eine einzige Gleichung mit mehr als einer unverzögerten endogenen Variablen

Da sich ergab, daß beim Auftreten verzögerter endogener Variablen, die mit den latenten Variablen stochastisch verbunden sind, der Schätzvorgang in vorläufig unüberschaubarer Weise beeinflußt wird[1], werden hier nur die Ergebnisse des Anwendungsfalls ohne verzögerte endogene Variablen resümiert.

Folgende Methoden erzielen dann identische Schätzergebnisse

(1) Die gewöhnliche Methode der kleinsten Quadrate liefert die gleichen Schätzergebnisse wie die verallgemeinerte Methode der kleinsten Quadrate, wenn die Varianz-Kovarianz-Matrix der latenten Variablen eine Skalarmatrix ist, d.h. wenn zugleich weder Autokorrelation noch Heteroskedastizität vorliegen. Allerdings weisen beide Methoden weder erwartungstreue noch konsistente Schätzungen auf, da nur eine der endogenen Variablen entsprechend ihren stochastischen Eigenschaften in den Schätzvorgang eingeht, die anderen endogenen Variablen jedoch unzulässigerweise als stochastisch unabhängig angesehen und behandelt werden.

(2) Die zweistufige Methode der kleinsten Quadrate und die dreistufige Methode der kleinsten Quadrate stimmen überein, da bei einer einzigen Gleichung die dritte Stufe nicht mehr angewandt werden kann. Beide Methoden liefern zumindest konsistente Schätzungen, da die restlichen endogenen Variablen durch Neuschätzung aufgrund der vorherbestimmten Variablen gewissermaßen entstochastiziert werden.

---

1) Insbesondere wenn zugleich Autokorrelation der latenten Variablen anzutreffen ist.

(3) Die indirekte Methode der kleinsten Quadrate und die Methode der größten Dichte bei beschränkter Information führen zu identischen Schätzergebnissen, wenn die reduzierte Form des Gesamtsystems von Strukturgleichungen, dem diese Gleichung angehören mag, berechnet werden kann und wenn die Strukturgleichungen exakt identifizierbar sind, damit eine Rückberechnung der Parameterschätzungen der Strukturgleichung gewährleistet ist. Sind diese beiden Bedingungen nicht zugleich erfüllt, so kann die indirekte Methode der kleinsten Quadrate nicht berechnet werden.

(4) Die Methode der größten Dichte bei vollständiger Information stimmt mit der Methode der größten Dichte bei beschränkter Information bzüglich der Schätzergebnisse überein[1].

Es ist festzuhalten, daß beide Methoden im betrachteten Anwendungsfall zumindest konsistente Schätzungen erzielen, da der stochastische Charakter aller endogenen Variablen der Gleichung in der Likelihoodfunktion zum Ausdruck kommt.

25.e. Ein interdependentes Gleichungssystem

Der Fall verzögerter endogener Variablen wurde wiederum ausgeschlossen, da sich sonst keine konkreten Ergebnisse

1) Es ist jedoch ausdrücklich zu betonen, daß eine solche Übereinstimmung auf den Eingleichungsfall zurückzuführen ist.

erzielen lassen. Auch ohne diesen Fall ergeben sich stark divergierende Schätzergebnisse bei einem Strukturgleichungssystem.

Die folgenden Ergebnisse wurden ermittelt:

(1) Die indirekte Methode der kleinsten Quadrate scheidet bei einem Strukturgleichungssystem als Schätzverfahren aus, da nicht nur die reduzierte Form berechenbar sein müßte, was im allgemeinen gelingen mag[1)], sondern auch jede einzelne Strukturgleichung exakt identifizierbar sein müßte, was bei größeren ökonometrischen Gleichungssystemen in der ökonomischen Realität kaum anzutreffen ist. Selbst wenn dieser einzigartige Fall vorliegen sollte, liefert die Methode der größten Dichte bei vollständiger Information effizientere Schätzungen.

(2) Die gewöhnliche Methode der kleinsten Quadrate liefert bei einem Strukturgleichungssystem nicht einmal konsistente Schätzungen, da die von mir für das Gleichungssystem als Ganzheit berechnete Verzerrung auch asymptotisch, d.h. mit zunehmendem Stichprobenumfang, nicht verschwinden wird. Diese Verzerrung ist noch stärker zu bewerten als die gleichungsweisen Verzerrungen aller Strukturgleichungen zusammen, da diese durch die Interdependenzen zwischen den einzelnen Strukturgleichungen verstärkt werden. Scheidet die indirekte Methode der

1) Wenn die zu Beginn dieser Synthese erwähnten fundamentalen Voraussetzungen vorliegen und die Parametermatrix der endogenen Variablen regulär ist und somit invertierbar ist.

kleinsten Quadrate schon von selbst aus, weil sie nicht berechnet werden kann, so muß hier festgestellt werden, daß die Schätzungen nach der gewöhnlichen Methode der kleinsten Quadrate immer dann berechnet werden können, wenn die sonstigen Schätzmethoden angewandt werden können. Die Simplizität der Berechnung darf aber nicht darüber hinwegtäuschen, daß die Schätzungen der Methode der kleinsten Quadrate bei einem Strukturgleichungssystem nicht konsistent sein können und daher nach Möglichkeit eines der anderen Verfahren vorgezogen werden sollte, wenn auch die Berechnung komplizierter und langwieriger ist.

(3) Die verallgemeinerte Methode der kleinsten Quadrate kann zwar wohlspezifizierte Autokorrelation und Heteroskedastizität als Schätzmethode berücksichtigen, diese Methode ist jedoch ungeeignet, wenn ein Strukturgleichungssystem geschätzt werden soll. Die verallgemeinerte Methode der kleinsten Quadrate erzielt in diesem Fall nicht einmal konsistente Schätzungen, da sie den gleichen Fehler inkorporiert wie die gewöhnliche Methode der kleinsten Quadrate: In jeder Gleichung wird nur eine der vorhandenen endogenen Variablen als endogene Variable angesehen und als solche behandelt, die restlichen endogenen Variablen der Gleichung werden unzulässigerweise als stochastisch unabhängig klassifiziert. Dieser Spezifikationsfehler führt zu der auch asymptotisch nicht verschwindenden Verzerrung. Wir erkennen, daß die verallgemeinerte Methode der kleinsten Quadrate zwar eine Verallgemeinerung der Methode der kleinsten Quadrate ist, daß diese Verallgemeinerung in bestimmten Fällen jedoch noch nicht zu einer spürbaren Verbesserung der Schätzergebnisse führt. Daß diese Methode bei einer Vereinigung mit den anderen Methoden doch zu spürbaren Fortschritten führen kann, werden wir später noch feststellen.

(4) Die zweistufige Methode der kleinsten Quadrate führt bei einem Strukturgleichungssystem zu konsistenten[1] Schätzungen. Diese Konsistenzeigenschaft ist allein auf die Entstochastizierung der restlichen endogenen Variablen jeder Gleichung durch Neuschätzung aufgrund der vorherbestimmten Variablen des Systems zurückzuführen. Allerdings ist die gleichungsweise Schätzung des Strukturgleichungssystems nicht so effizient wie die Simultanschätzung nach der dreistufigen Methode.

(5) Die dreistufige Methode der kleinsten Quadrate führt zu konsistenten Schätzungen, die wegen ihrer Effizienz im Vergleich zur zweistufigen Methode deren Schätzungen vorzuziehen ist. Diese Effizienz relativ zur zweistufigen Methode ist darauf zurückzuführen, daß die dreistufige Methode explizit die Streuungs-Kovarianz-Matrix der latenten Variablen aller Gleichungen berücksichtigt, allerdings nur für den Spezialfall, daß die Streuungs-Kovarianz-Matrizen der einzelnen Strukturgleichungen Skalarmatrizen sein müssen. Also sind für die einzelnen Strukturgleichungen Autokorrelation und zugleich Heteroskedastizität nicht zulässig. Da aber nicht eine Identität aller Skalarmatrizen (Steuungs-Kovarianz-Matrizen der latenten Variablen) der einzelnen Gleichungen gefordert wird, wie dies bei der gleichungsweisen Anwendung der zweistufigen Methode implizit vorausgesetzt wird, ist die dreistufige Methode allgemeiner und effizienter als die zweistufige Methode.

(6) Die Methode der größten Dichte bei beschränkter Information führt bei einem Strukturgleichungssystem zu konsistenten Schätzungen. Dies ist das Ergebnis der

---

1) Wohlgemerkt, nur solange keine verzögerten endogenen Variablen auftreten. Das gilt auch für die nachfolgende dreistufige Methode.

korrekten Spezifikation aller endogenen Variablen des Systems als stochastisch abhängige Variablen, die sich in der Dichtefunktion niederschlägt. Die gleichungsweise oder untersystemweise Anwendung der Methode der größten Dichte bei beschränkter Information ist jedoch im Vergleich zur Simultanschätzung der Methode der größten Dichte bei vollständiger Information ineffizienter.

(7) Die Methode der größten Dichte bei vollständiger Information führt zu konsistenten Schätzungen, die im Vergleich zur Methode der größten Dichte bei beschränkter Information effizienter sind. Diese Effizienz ist darauf zurückzuführen, daß die in den a-priori-Restriktionen der jeweils von der Methode der größten Dichte bei beschränkter Information nicht simultan berücksichtigten Gleichungen enthaltenen Informationen bei der letzteren Methode verloren gehen. Diese Informationen werden nicht wieder gewonnen, wenn die Methode der größten Dichte bei beschränkter Information zu diesen Gleichungen gelangt, da dann wieder die a-priori-Restriktionen der ersteren Gleichungen vernachlässigt werden. Dieser Punkt veranschaulicht den Vorteil simultaner Verfahren. Von höchstem Interesse ist nun der Vergleich der Methode der größten Dichte mit dem anderen Simultanverfahren, nämlich der dreistufigen Methode der kleinsten Quadrate.

An dieser Stelle muß darauf hingewiesen werden, daß die Methode der größten Dichte bei vollständiger Information zwar leichter zu berechnen ist, wenn die Varianz-Kovarianz-Matrizen der latenten Variablen der

einzelnen Strukturgleichungen Skalarmatrizen[1] sind, dies aber keine restriktive Voraussetzung für die Anwendung des Verfahrens ist. Dagegen kann die dreistufige Methode der kleinsten Quadrate dann und nur dann berechnet werden, wenn die Varianz-Kovarianz-Matrizen der latenten Variablen der einzelnen Strukturgleichungen Skalarmatrizen sind.

Wir sehen, daß die Voraussetzungen der dreistufigen Methode der kleinsten Quadrate restriktiver sind als die Voraussetzungen der Methode der größten Dichte bei vollständiger Information. Damit wird offenbar, daß die Methode der größten Dichte bei vollständiger Information effizienter ist als die dreistufige Methode der kleinsten Quadrate. Wird die dreistufige Methode weiter ausgebaut, wie im folgenden beabsichtigt ist, so läßt sich dieser Mangel mildern.

1) Innerhalb der einzelnen Strukturgleichungen sind dann Autokorrelationen und Heteroskedastizität (über die Zeit variierende Streuungen) ausgeschlossen.

26. Allgemeine Probleme der Anwendbarkeit ökonometrischer Schätzverfahren auf ökonomische Tatbestände

Das Hauptanliegen bei der Gründung der Econometric Society und der Begründung der Disziplin Ökonometrie bestand in der Koordination der Bereiche Mathematik, Statistik und Ökonomie und insbesondere in der Anwendung mathematisch-statistischer Verfahren auf ökonomische Probleme.

Wenn man heute kritische Stellungnahmen, wie z.B. die von Herman Wold[1] oder Clifford Hildreth[2], liest, kann man sehr leicht zu der Ansicht gelangen, daß sich die weitere Entwicklung ökonometrischer Schätzverfahren nicht lohnt, da einmal die Ökonomie unkompliziert ist und Interdependenzen keine wesentliche Rolle spielen[3] und zum anderen die "mathematischen Komplikationen"[4], die bei der Berücksichtigung von Gleichungssystemen mit Interdependenzen auftreten, von der Benutzung komplizierter Verfahren abraten lassen[5]. Außerdem wird von diesen Kritikern[6] festgestellt, daß kompliziertere Verfahren nur in geringem Maße zu empirischen Untersuchungen herangezogen wurden, und nur wenige

---

1) Wold, H. (in Zusammenarbeit mit Juréen, L.): a.a.O., S. 48-53.

2) Hildreth, C.: Simultaneous equations: Any verdict yet? Econometrica, 28(1960), S. 846-854.

3) Wold, H. (in Zusammenarbeit mit Juréen, L.): a.a.O., S. 53.

4) Hildreth: a.a.O., S. 848.

5) Eine symptomatische Stellungnahme der empirischen Kritiker ist die von F.V. Waugh: The place of least squares in econometrics. Econometrica, 29(1961), S. 386-396. Abgesehen von einer überaus erfolgsorientierten Argumentation (Mausefallenbeispiel auf S. 387), wertet Waugh Diagramme und Einfachregressionen (S. 395) höher als mit dem Computer erzielte Ergebnisse, wobei Waugh offen läßt, ob letztere Ergebnisse nach der Methode der kleinsten Quadrate berechnet wurden.

6) Wold, H.: (in Zusammenarbeit mit Juréen, L.): a.a.O., S. 15. Hildreth: a.a.O., S. 847.

ökonomische Anwendungsmöglichkeiten für Schätzverfahren bestehen, die alle Gleichungen eines Gleichungssystems simultan schätzen.

Abgesehen davon, daß der letztere Einwand auch die dreistufige Methode der kleinsten Quadrate treffen würde, läßt sich aus der infolge mathematischer Komplexität geringen bisherigen empirischen Verwertung eines Verfahrens noch nichts über seine potentiellen Erklärungsmöglichkeiten und seine empirische Brauchbarkeit aussagen.

Zum anderen muß festgestellt werden, daß Wold's "Demand Analysis" zwar eine ökonometrische Studie, jedoch spezifisch mikroökonomischen Charakters ist, die sich die Untersuchung einzelner Märkte und Haushaltsbefragungen[1] zum Ziel setzte. In diesem Rahmen können Simultaneitäten durch die Annahme verzögerter Anpassungen der Wirtschaftssubjekte ausgeschaltet werden.

Dagegen muß bei der Untersuchung einer Gesamtwirtschaft mit Hilfe eines makroökonomischen Modells in aller Regel mit Interdependenzen der ökonomischen Variablen gerechnet werden. Diese Interdependenzen[2] können das Ergebnis realer Abläufe sein, sie können jedoch auch das Ergebnis der vorausgehenden Aggregationsprozesse oder der materiell-statistischen Besonderheiten der Daten sein.

---

1) Vgl. hierzu auch Klein, L.R.: Single equation versus equation system methods of estimation in econometrics. Econometrica, 28(1960), S. 866-871, s. bes. S. 871.

2) Vgl. Menges (1961): a.a.O., S. 55-56.

27. Stellungnahme zu den Monte-Carlo-Studien[1)]

In neuester Zeit ist eine stärkere Abwendung von den klassischen, zumeist asymptotischen Schätzkriterien von R.A. Fisher zu Schätzkriterien für kleine Stichproben zu beobachten. Diese Entwicklung begann mit der kritischen Erkenntnis, daß ein Schätzverfahren wohl hervorragende asymptotische Eigenschaften besitzen könne, aber bei kleinen Stichproben schlechtere Ergebnisse erziele als ein anderes Schätzverfahren mit nicht hervorstechenden asymptotischen Eigenschaften. Diese Erkenntnis ist insbesondere gravierend, da in den Sozialwissenschaften Zeitreihen mit nur etwa 30 Beobachtungen eine Seltenheit darstellen und zumeist durch die Gefahr von Strukturbrücken aufgrund der verschiedensten Einflüsse mit Vorsicht zu verwenden sind. Infolgedessen wurden sogenannte Monte-Carlo-Simulationen[2)] durchgeführt, bei denen eine Grundgesamtheit mit Hilfe von Zufallszahlen und einem hypothetischen Modell von Strukturgleichungen

---

1) Vgl. die Darstellung der Monte-Carlo-Studien durch Johnston in Johnston: a.a.O., S. 275-295 sowie folgende Monte-Carlo-Untersuchungen:

Ladd, G.W.: Effects of shocks and errors in estimation: An empirical comparison. Journal of Farm Economics, 38(1956), S. 485-495.

Neiswanger, W.A. u. Yancey, T.A.: Parameter estimates and autonomous growth. Journal of the American Statistical Association, 54(1959), S. 389-402.

Wagner, H.: A Monte Carlo study of estimates of simultaneous linear structural equations. Econometrica, 26(1958), S. 117-133.

Nagar, A.L.: A Monte Carlo study of alternative simultaneous equation estimators. Econometrica, 28(1960), S. 573-590.

Summers, R.: A capital intensive approach to the small sample properties of various simultaneous equation estimators. Econometrica, 33(1965), S. 1-41.

2) Vgl. hierzu Christ, C.F.: Simultaneous equation estimation: Any verdict yet? Econometrica, 28(1960), S. 835-845, s. bes. S. 853.

mit vorgegebenen Parametern generiert wird. Aus dieser "wahren" Grundgesamtheit werden dann in üblicher Weise Stichproben der "Beobachtungswerte" der endogenen und vorherbestimmten Variablen gezogen, wobei die - jetzt gewissermaßen vergessenen - Parameter geschätzt und mit den "wahren" Parametern verglichen werden. Interessant ist, daß anstelle großer Stichproben nun eine große Anzahl von Stichproben mit kleinem Stichprobenumfang gezogen wird, was gewissermaßen den Verlust der Berechnung asymptotischer Eigenschaften ausgleichen soll.

An dieser Stelle möchte ich zwei kritische Punkte zu Monte-Carlo-Experimenten festhalten:

Einmal weisen gewisse Schätzverfahren asymptotische Eigenschaften wie z.B. die Konsistenz als Minimalforderung[1] auf, weil sie die Struktur des Schätzproblems korrekt erfassen: So ist z.B. die Methode der größten Dichte im Gegensatz zur gewöhnlichen Methode der kleinsten Quadrate bei mehreren endogenen Variablen einer Gleichung konsistent, weil sie die nicht ausgewählten endogenen Variablen nicht als stochastisch unabhängig behandelt. Diese bessere Erkenntnis des Schätzproblems wird sich auch bei kleinen Stichproben schwerlich als Hindernis erweisen. Wenn sich also bei Monte-Carlo-Studien eine andere Hierarchie der Schätzverfahren nach ihren Schätzqualitäten ergeben sollte, so kann dies nur an speziellen Eigenheiten des gewählten Modells liegen. Wir wollen dies gerade anhand der Konsistenz untersuchen. Die Konsistenz besagt nichts anderes, als die Schwankung der Parameterschätzung um den festen "wahren" Parameter mit gegen Unendlich strebenden Stichprobenumfang konvergiert. So kann bei kleinen Stichproben die Schwankung

1) Vgl. Menges (1961): a.a.O., S. 89.

(etwa als Oszillation) der Parameterschätzung um den "wahren" Parameter zwar beträchtlich sein, sie wird jedoch bei Konsistenz des Verfahrens mit zunehmendem Stichpobenumfang immer stärker abnehmen. Da man in den Sozialwissenschaften zwar nur kleine Stichproben ziehen kann, gleichwohl die Stichprobentätigkeit als solche weit verbreitet ist, erscheint es nicht vorteilhaft, z.B. ein nicht konsistentes Verfahren vorzuziehen, bei dem die obige Schwankung zufälligerweise einmal geringer war als bei dem konsistenten Verfahren. Vielmehr ist zu vermuten, daß die große Anfangsschwankung durch die asymptotische Verzerrung des nicht konsistenten Verfahrens zum Teil kompensiert wurde. Würde der Stichprobenumfang nur geringfügig ausgedehnt, so würden sich beide Fehlerquellen bei einer Oszillation eher in einer stärkeren Schwankung des nicht konsistenten Verfahrens im Vergleich zum konsistenten Verfahren auswirken.

Der zweite kritische Punkt bezieht sich eher auf das Vorgehen der Monte-Carlo-Simulation. Soweit ich beurteilen kann, wurden bisher die hypothetischen Modelle mit ihren Parametern nicht willkürlich algebraisch vorgegeben, sondern basieren auf bereits durchgeführten empirischen Schätzungen, welche ihrerseits stichprobenfehlerbehaftet sein werden. Zudem weisen diese empirischen Modelle zumeist starke a-priori-Restriktionen auf, die den Generalitätsgrad solcher Modelle sehr einengen. Da zumeist das Interesse auf eine Gleichung, etwa die Konsumfunktion konzentriert wird, werden zugleich die Anwendungsmöglichkeiten von Gleichungssystem-Schätzverfahren nicht durchgespielt. Infolgedessen haben die meisten bisherigen Monte-Carlo-Simulationen lediglich einen anschaulichen Wert.

Überzeugender können Monte-Carlo-Studien werden, wenn das zugrundegelegte Modell allgemein oder für mehrere ökonomische Anwendungsfälle[1] formuliert ist.

Zumindest sollte auch das Modell nicht fest vorgeben, sondern in irgendeiner Weise simuliert werden. Da dies die Simulation erheblich erschwert, ist vorerst nicht mit solchen, auf allgemeinerer Grundlage stehenden Monte-Carlo-Ergebnissen zu rechnen. Solange sind indes die Ergebnisse von Monte-Carlo-Studien auch im Hinblick auf die asymptotischen Eigenschaften kritisch zu betrachten.

1) Christ zieht folgende parallele Schlußfolgerung: "The important task ahead is to learn more about how to decide which estimation method is likely to be best for any given actual econometric problem". Christ: a.a.O., S. 845.

28. Probleme beim Auftreten von Autokorrelation[1)]

28.a. Eine einzige Gleichung mit nur einer unverzögerten endogenen Variablen.

Während im allgemeinen bei ökonometrischen Schätzverfahren angenommen wird, daß der Wert der latenten Variablen[2)] einer Gleichung zu einem bestimmten Zeitpunkt von ihrem Wert zu einem anderen (früheren oder späteren) Zeitpunkt unabhängig ist, wird dies in der Praxis meist nicht der Fall sein. Die zeitliche Abhängigkeit der Werte der latenten Variablen, welche als Autokorrelation bekannt ist, kann etwa durch folgenden autoregressiven Markoff-Prozeß 1. Ordnung gekennzeichnet sein.[3)]

$$(28.1) \qquad u_t = \rho\, u_{t-1} + \eta_t, \qquad \text{mit } |\rho| < 1,$$

wobei die üblichen Voraussetzungen (Erwartungswert Null, Diagonalmatrix mit Homoskedastizität) nunmehr für die Abweichung η_t und nicht mehr für die latente Variable u_t selbst gelten. Es läßt sich dann die Streuungs-Kovarianz-Matrix der latenten Variablen angeben mit[4)]

$$(28.2) \quad V = E(u\,u') = \sigma_u^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix}$$

---

1) Vgl. Johnston: a.a.O., S. 177-200.

2) Wir müssen uns die - nicht beobachtbare - latente Variable als eine in der Zeit veränderliche Variable vorstellen.

3) Johnston: a.a.O., S. 178.

4) Johnston: a.a.O., S. 180.

wegen $E(u_t\ u_{t-1}) = \rho^1\ \sigma_u^2$

und wobei u der Spaltenvektor der latenten Variablen bei n Perioden ist.

Ist diese Spezifikation insgesamt zutreffend, so führt die Aitkensche verallgemeinerte Methode der kleinsten Quadrate im Vergleich zur gewöhnlichen Methode der kleinsten Quadrate zu effizienten Schätzungen.

28.b. Eine einzige Gleichung mit mehreren unverzögerten endogenen Variablen

Treten mehrere unverzögerte endogene Variablen in einer Gleichung auf, so erscheint u.a. die Verwendung der zweistufigen Methode der kleinsten Quadrate als konsequent. Wird darüberhinaus Autokorrelation vermutet, so bietet sich die von Theil[1] entwickelte verallgemeinerte zweistufige Methode der kleinsten Quadrate an. Diese Methode ist die Verbindung der verallgemeinerten Methode der kleinsten Quadrate, welche die Streuungs-Kovarianz-Matrix der latenten Variablen explizit verwendet, mit der zweistufigen Methode der kleinsten Quadrate, welche insbesondere die restlichen endogenen Variablen neuschätzt. Für dieses Spezialproblem ist also zu erwarten, daß die verallgemeinerte zweistufige Methode der kleinsten Quadrate im Vergleich zu anderen Schätzverfahren, die nicht mehr Informationen verwerten, effizient ist.

---

1) Theil (1961): a.a.O., S. 344-345.

28.c. Ein interdependentes Gleichungssystem im Zusammenhang mit Autokorrelation.

Rothenberg und Leenders[1] kritisieren, daß die dreistufige Methode der kleinsten Quadrate nicht mehr effizient sein kann, wenn die Streuungs-Kovarianz-Matrix a-priori-Restriktionen unterworfen ist, solange in mindestens einer Gleichung mehr als eine unverzögerte endogene Variable auftritt. An dieser Stelle muß berücksichtigt werden, daß die dreistufige Methode der kleinsten Quadrate für jede Strukturgleichung voraussetzt, daß keine Autokorrelation der latenten Variablen dieser Gleichung besteht. Dagegen gestattet die dreistufige Methode der kleinsten Quadrate zu jedem fortgesetzten Zeitpunkt eine gewissermaßen "gleichzeitige" Kovarianz der latenten Variablen verschiedener Strukturgleichungen, die aber zeitlich konstant sein muß. Daraus resultiert, daß die Kovarianz der latenten Variablen der i-ten Strukturgleichung zum Zeitpunkt t mit der latenten Variablen der j-ten Strukturgleichung ($i \neq j$) zum Zeitpunkt t' ($t' \neq t$) Null sein muß, damit die dreistufige Methode der kleinsten Quadrate angewandt werden kann.

Da ich in den Schriften von H. Theil bisher vergeblich nach einer Methode zur Lösung dieses Problems gesucht habe, möchte ich nunmehr gewissermaßen eine verallgemeinerte dreistufige Methode der kleinsten Quadrate entwickeln. Diese ist die Verbindung der verallgemeinerten Methode der kleinsten Quadrate von Aitken und der dreistufigen Methode der kleinsten Quadrate von Zellner und Theil. Da diese Methode die Restriktionen der Streuungs-Kovarianz-Matrix der latenten Variablen explizit berücksichtigt, wird

1) Rothenberg, T.J. u. Leenders, C.T.: Efficient estimation of simultaneous equation systems. Econometrica, 32(1964), S. 57-81, s. bes. S. 72.

die Kritik von Rothenberg und Leenders inhaltslos, so daß anzunehmen ist, daß die verallgemeinerte dreistufige Methode der kleinsten Quadrate der Methode der größten Dichte bei vollständiger Information nicht länger an Effizienz nachsteht.

Im Gegensatz zur dreistufigen Methode ersetzen wir bei der verallgemeinerten dreistufigen Methode der kleinsten Quadrate die Voraussetzung,

"daß die latenten Variablen der Strukturgleichungen den Mittelwert Null haben, zeitlich unabhängig voneinander sind und homoskedastisch in dem Sinne sind, daß die Streuungen der latenten Variablen der einzelnen Gleichungen und die Kovarianzen der latenten Variablen zu gegebenem Zeitpunkt der Gleichungen miteinander endlich und zeitlich invariant sind."[1)]

durch die allgemeinere Voraussetzung:
Die latenten Variablen der Strukturgleichungen haben den Mittelwert (Erwartungswert) Null und eine endliche reguläre Varianz-Kovarianz-Matrix.

Dann erhalten wir für die Streuungs-Kovarianz-Matrix V der latenten Variablen aller G Gleichungen

$$(28.3) \qquad V = \begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1G} \\ V_{21} & V_{22} & \cdots & V_{2G} \\ \vdots & \vdots & & \vdots \\ V_{G1} & V_{G2} & \cdots & V_{GG} \end{bmatrix}$$

1) Vgl. Zellner u. Theil: a.a.O., S. 55. Übersetzung des Verfassers dieser Arbeit, wobei die Terminologie dieser Arbeit verwendet wurde.

mit

$$(28.4)\quad V_{ij} = E \begin{bmatrix} (u_{i1}u_{j1}) & (u_{i1}u_{j2}) & \dots & (u_{i1}u_{jn}) \\ (u_{i2}u_{j1}) & (u_{i2}u_{j2}) & \dots & (u_{i2}u_{jn}) \\ \vdots & \vdots & & \vdots \\ (u_{in}u_{j1}) & (u_{in}u_{j2}) & \dots & (u_{in}u_{jn}) \end{bmatrix}$$

für (i,j=1,...,G),

wenn n die Zahl der Beobachtungsperioden ist.

Die Skalarmatrix $V_{ij} = \sigma_{ij} I_n$ ist nunmehr nur noch ein Spezialfall von (28.4). Im allgemeinen wird also die Streuungs-Kovarianz-Matrix der latenten Variablen der i-ten mit der j-ten Gleichung auch oberhalb und unterhalb der Hauptdiagonalen von Null verschiedene Werte aufweisen, wenn Autokorrelation vorliegt. Dies sieht man besonders an der Streuungs-Kovarianz-Matrix V_{ii} für (i=1,...,G), die ein Element der Hauptdiagonalen von (28.3) ist. Des weiteren muß nicht mehr Homoskedastizität vorliegen, was man ebenfalls deutlich bei der Matrix V_{ii} für (i=1,...,G) erkennt, da die Hauptdiagonalelemente nicht mehr identisch sein müssen.

Im Schätzvorgang unterscheidet sich die verallgemeinerte dreistufige Methode grundsätzlich nicht von der dreistufigen Methode der kleinsten Quadrate. Hier wie dort[1] wird das in Simultanform geschriebene System

$$(28.5)\quad \begin{bmatrix} X'y_1 \\ X'y_2 \\ \vdots \\ X'y_G \end{bmatrix} = \begin{bmatrix} X'Z_1 & 0 & \dots & 0 \\ 0 & X'Z_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & X'Z_G \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_G \end{bmatrix} + \begin{bmatrix} X'u_1 \\ X'u_2 \\ \vdots \\ X'u_G \end{bmatrix}$$

1) Vgl. Zellner u. Theil: a.a.O., S. 57-58.

nach der verallgemeinerten Methode der kleinsten Quadrate geschätzt.

Der wesentliche Unterschied besteht nur in der Spezifikation der Streuungs-Kovarianz-Matrizen V_{ij} für (i,j=1,...,G). Bei der verallgemeinerten dreistufigen Methode muß nicht jede dieser Matrizen als Skalarmatrix spezifiziert werden wie bei der dreistufigen Methode der kleinsten Quadrate. Dies ist nicht nur als Vorteil zu werten. Kritisch muß angefügt werden, daß die Spezifikation der Streuungs-Kovarianz-Matrix der latenten Variablen den Schätzvorgang wesentlich beeinflußt und daß durch eine Fehlspezifikation das Verfahren in Frage gestellt ist. Dieser Einwand trifft allerdings auch die bekannte dreistufige Methode der kleinsten Quadrate, da die Annahme einer Skalarmatrix selbst eine Fehlspezifikation sein kann, wenn Autokorrelation oder Heteroskedastizität auftreten.

29. Schlußbetrachtung

Diese Arbeit sollte durch einen kritischen Vergleich der wichtigsten zur Zeit bekannten ökonometrischen Schätzverfahren deren relative Vorzüge und Nachteile herausarbeiten. Durch Konfrontation dieser Schätzverfahren mit den verschiedenen Typen ökonometrischer Modelle und mit diversen ökonometrischen Problemen wurden einige Ansatzpunkte zur Entwicklung neuerer Schätzverfahren gefunden. Diese neuen Schätzverfahren ergeben sich im wesentlichen durch die Kombination bisheriger Schätzverfahren, welche zwar verschiedene Probleme - zumeist jedoch nur für einen Spezialfall - lösen können. Es erscheint daher nicht abwegig, wenn die in der Realität zu beobachtende Vermischung ökonometrischer Phänomene durch mehrphasige Schätzverfahren erfaßt wird.

Die Entwicklung weiterer mehrphasiger Schätzverfahren ist jedoch nicht zuletzt von der Entwicklung der Computer-Technik abhängig, da die Überprüfung mehr oder minder komplizierter Verfahren durch empirisches Arbeiten mit Hilfe schnellrechnender Computer auf die Dauer die Skepsis gegen mathematische Methoden mindern wird und Anregungen für weitere Forschungen geben kann.

LITERATURVERZEICHNIS

Aitken, A.C.: On least squares and linear combination of observations. Proceedings of the Royal Society of Edinburgh, 55(1934-1935), S. 42-48.

Aitken, A.C.: Determinants and Matrices. Edinburgh-London 1939 (9. Aufl. 1958).

Anderson, T.W.: Estimation of the parameters of a single equation by the limited-information maximum-likelihood method. In: Statistical Inference in Dynamic Economic Models. Hrsg.: Tjalling C. Koopmans. New York-London 1950, S. 311-322.

Anderson, T.W. u. Rubin, H.: Estimation of the parameters of a single equation in a complete system of stochastic equations. The Annals of Mathematical Statistics, 20(1949), S. 46-63.

Anderson, T.W. u. Rubin, H.: The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations. The Annals of Mathematical Statistics, 21(1950), S. 570-582.

Chernoff, H. u. Divinsky, N.: The computation of maximum-likelihood estimates of linear structural equations. In: Studies in Econometric Method. Hrsg.: W.C. Hood u. Tjalling C. Koopmans. New York-London 1953, S. 263-269.

Chernoff, H. u. Rubin, H.: Asymptotic properties of limited-information estimates under generalized conditions. In: Studies in Econometric Method. Hrsg.: W.C. Hood und Tjalling C. Koopmans. New York-London 1953, S. 200-212.

Chow, G.C.: A comparison of alternative estimators for simultaneous equations. Econometrica, 32(1964), S. 532-553.

Christ, C.F.: Simultaneous equation estimation: Any verdict yet? Econometrica, 28(1960), S. 835-845.

Cramér, H.: Mathematical Methods of Statistics. Princeton 1946.

Eisenhart, C.: The background and evolution of the method of least squares. Digest prepared for distribution to participants of the 34th Session of the International Statistical Institute, Ottawa, Canada, 21-29 August 1963.

Fisher, R.A.: On an absolute criterion for fitting frequency curves. Messenger of Mathematics, 41(1912), S. 155 ff.

Fisher, R.A.: On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society, Series A, 222(1922), S. 309-368. (Wiederabdruck in: R.A. Fisher: Contributions to Mathematical Statistics. New York-London 1950, Paper 10).

Fisz, M.: Wahrscheinlichkeitsrechnung und mathematische Statistik. 4. Aufl., Berlin 1966.

Haavelmo, T.: The statistical implications of a system of simultaneous equations. Econometrica, 11(1943), S. 1-12.

Hildreth, C.: Simultaneous equations: Any verdict yet? Econometrica, 28(1960), S. 846-854.

Johnston, J.: Econometric Methods. New York-San Francisco -Toronto-London 1963.

Kendall, M.G. u. Stuart, A.: The Advanced Theory of Statistics. Bd. 2: Inference and Relationship. London 1961.

Klein, L.R.: Single equation versus equation system methods of estimation in econometrics. Econometrica, 28(1960), S. 866-871.

Koopmans, T.C. u. Hood, W.C.: The estimation of simultaneous linear economic relationships. In: Studies in Econometric Method. Hrsg.: W.C. Hood u. Tjalling C. Koopmans. New York 1953, S. 112-199.

Koopmans, T.C., Rubin, H. u. Leipnik, R.B.: Measuring the equation systems of dynamic economics. In: Statistical Inference in Dynamic Economic Models. Hrsg.: Tjalling C. Koopmans. New York-London 1950, S. 53-237.

Ladd, G.W.: Effects of shocks and errors in estimation: An empirical comparison. Journal of Farm Economics, 38(1956), S. 485-495.

Lüdeke, D.: Schätzprobleme in der Ökonometrie. Würzburg 1964.

Mann, H.B. u. Wald, A.: On the statistical treatment of linear stochastic difference equations. Econometrica, 11(1943), S. 173-220.

Menges, G.: Die Unterschiede in den Erkenntnisweisen von Statistik und Ökonometrie. Annales Universitatis Saraviensis (Rechts- und Wirtschaftswissenschaften), 7(1959), S. 191-202.

Menges, G.: Ökonometrie. Wiesbaden 1961.

Menges, G.: Three essays in econometrics. Statistische Hefte, 4(1963), S. 1-37.

Menges, G.: Die ökonometrische Struktur und die Frage ihrer Konstanz. In: Strukturwandlungen einer wachsenden Wirtschaft. Bd. 2, Berlin 1964 (Schriften des Vereins für Socialpolitik, Neue Folge, Bd. 30), S. 989-1004.

Menges, G.: Grundriß der Statistik. Teil 1: Theorie. Köln-Opladen 1968.

Menges, G. u. Diehl, H.: Das Stabilitätsproblem in der Ökonometrie. Statistische Hefte, 6(1965), S. 27-52.

Nagar, A.L.: The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. Econometrica, 27(1959), S. 575-595.

Nagar, A.L.: A Monte Carlo study of alternative simultaneous equation estimators. Econometrica, 28(1960), S. 573-590.

Neiswanger, W.A. u. Yancey, T.A.: Parameter estimates and autonomous growth. Journal of the American Statistical Association, 54(1959), S. 389-402.

Pearson, K.: Contributions to the mathematical theory of evolution. I. Philosophical Transactions of the Royal Society, Series A, 185(1894), S. 71 ff.

Pearson, K.: Contributions to the mathematical theory of evolution. IV.: On the probable errors of frequency constants and on the influence of random selection on variation and correlation. Philosophical Transactions of the Royal Society, Series A, 191(1898), S. 229 ff.

Pearson, K.: On the systematic fitting of curves to observations and measurements. Biometrika, 1(1901/1902), S. 265-303 und Biometrika, 2(1902/1903), S. 1-23.

Rothenberg, T.J. u. Leenders, C.T.: Efficient estimation of simultaneous equation systems. Econometrica, 32(1964), S. 57-81.

Rubin, H.: Consistency of maximum-likelihood estimates in the explosive case. In: Statistical Inference in Dynamic Economic Models. Hrsg.: Tjalling C. Koopmans. New York-London 1950, S. 356-364.

Schneeweiß, H.: Vorlesung Ökonometrie, SS 1966, an der Universität des Saarlandes (hektographiert).

Strotz, R.H. u. Wold, H.: Recursive versus nonrecursive systems: An attempt at synthesis. Econometrica, 28 (1960), S. 417-427.

Summers, R.: A capital intensive approach to the small sample properties of various simultaneous equation estimators. Econometrica, 33(1965), S. 1-41.

Taylor, L.D. u. Wilson, R.A.: Three-Pass Least Squares: A method for estimating models with a lagged dependent variable. The Review of Economics and Statistics, 46(1964), S. 329-346.

Theil, H.: The aggregation implications of identifiable structural macrorelations. Econometrica, 27(1959), S. 14-29.

Theil, H.: Economic Forecasts and Policy. 2. Aufl., Amsterdam 1961 (1. Aufl.: Amsterdam 1958).

Valvanis, S.: Econometrics. An Introduction to Maximum Likelihood Methods. New York-Toronto-London 1959 (posthum hrsgg. von A.H. Conrad).

Wagner, H.: A Monte Carlo study of estimates of simultaneous linear structural equations. Econometrica, 26(1958), S. 117-133.

Waugh, F.V.: The place of least squares in econometrics. Econometrica, 29(1961), S. 386-396.

Wold, H.: A generalization of causal chain models. Econometrica, 28(1960), S. 443-463.

Wold, H.: Forecasting by the chain principle. In: Econometric Model Building. Hrsg.: H.O.A. Wold. Amsterdam 1964, S. 5-36.

Wold, H. (in Zusammenarbeit mit Juréen, L.): Demand Analysis. New York-Stockholm 1953.

Zellner, A. u. Theil, H.: Three-stage least squares: Simultaneous estimation of simultaneous equations. Econometrica, 30(1962), S. 54-78.

Zurmühl, R.: Matrizen und ihre technischen Anwendungen. Berlin-Göttingen-Heidelberg 1964.

difo-druck · schmacht bamberg